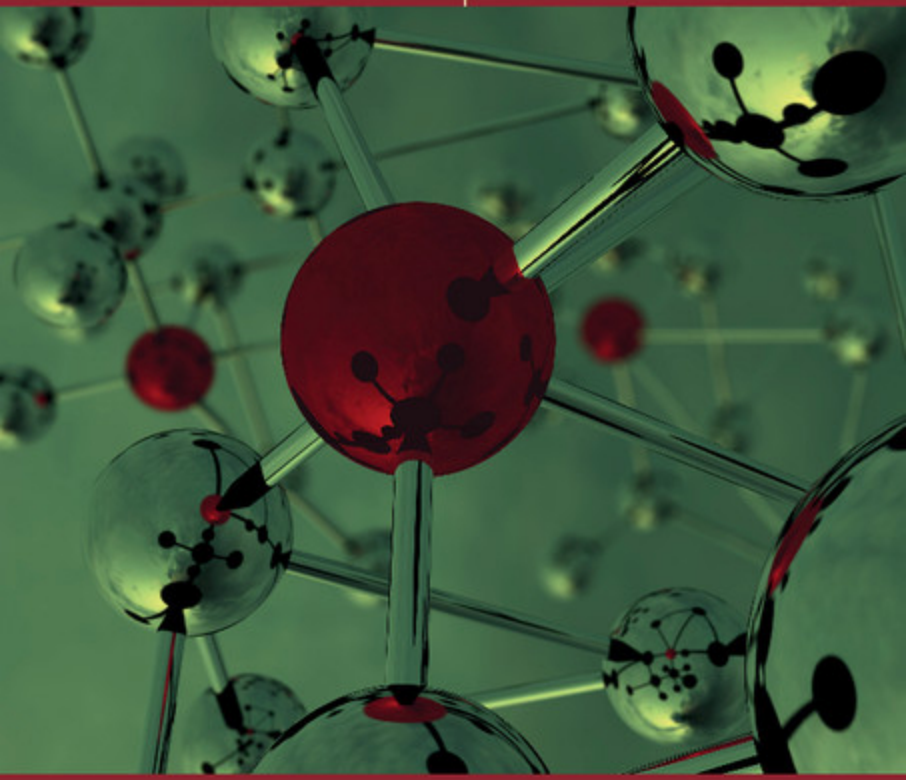
CAMBRIDGE INTRODUCTIONS TO PHILOSOPHY

An Introduction to the Philosophy of Science

KENT W. STALEY



An Introduction to the Philosophy of Science

This book guides readers by gradual steps through the central concepts and debates in the philosophy of science. Using concrete examples from the history of science, Kent W. Staley shows how seemingly abstract philosophical issues are relevant to important aspects of scientific practice. Structured in two parts, the book first tackles the central concepts of the philosophy of science, such as the problem of induction, falsificationism, and underdetermination, and important figures and movements, such as the logical empiricists, Thomas Kuhn, and Paul Feyerabend. The second part turns to contemporary debates in the philosophy of science, such as scientific realism, explanation, the role of values in science, the different views of scientific inference, and probability. This broad yet detailed overview will give readers a strong grounding whilst also providing opportunities for further exploration. It will be of particular interest to students of philosophy, the philosophy of science, and science.

KENT W. STALEY is Associate Professor of Philosophy at Saint Louis University, Missouri. His publications include *The Evidence for the Top Quark: Objectivity and Bias in Collaborative Experimentation* (Cambridge, 2004).

An Introduction to the Philosophy of Science

KENT W. STALEY
Saint Louis University, Missouri

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521129992

© Kent W. Staley 2014

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2014

Printed in the United Kingdom by Clays, St Ives plc

A catalog record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Staley, Kent W. (Kent Wade), 1963– author.
An introduction to the philosophy of science / Kent W. Staley.
pages cm
ISBN 978-0-521-11249-9 (Hardback) – ISBN 978-0-521-12999-2 (Paperback)
1. Science–Philosophy. I. Title.
Q175.S735 2014
501–dc23 2014014318

ISBN 978-0-521-11249-9 Hardback
ISBN 978-0-521-12999-2 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

To Peter Achinstein and the late Paul C. L. Tang, who introduced me to philosophy of science.

Contents

List of figures *page* ix
List of tables x
Preface: philosophy of science for philosophers, scientists, and everyone else xi
Acknowledgments xvi
List of abbreviations xvii

Part I Background and Basic Concepts **1**

1 Some problems of induction 3

2 Falsificationism: science without induction? 15

3 Underdetermination 26

4 Logical empiricism and scientific theories 39

5 Kuhn: scientific revolutions as paradigm changes 55

6 Lakatos: scientific research programs 71

7 Feyerabend: epistemological anarchism 85

Part II Ongoing Investigations **103**

8 Reasoning with probability: Bayesianism 105

9 Reasoning with probability: frequentism 135

10 Realism and anti-realism 167

11 Explanation 199

12 Values in science 229

References 258
Index 276

Figures

1.1 Newton's first experimental arrangement. 9
1.2 Newton's two-prism experimental arrangement. 10
3.1 Foucault's experimental arrangement. 29
5.1 Three phylogenetic trees of the sawfly (genus *Neodiprion*). 65
9.1 The binomial distribution for Y_i ($p = 1/6$, $N = 36$). 139
9.2 Dice game results, $N = 100$ 139
10.1 A cloud chamber photograph of a track left by a positron. 170
11.1 A schematic drawing of an action potential. 203

Tables

2.1 Doughnuts and other things. 17
4.1 Structure $\mathfrak{A}$: A model of L_1. 52
4.2 Structure $\mathfrak{B}$: A model of L_2. 53
8.1 A sure-loss contract for Samiksha. 115
9.1 Data from the rolling of a die. 137
9.2 Shopping list for a Fisherian test. 142

Preface: philosophy of science for philosophers, scientists, and everyone else

For better and for worse, in ways both obvious and subtle, the work of scientists has helped to shape the world around us. The obvious impacts of science on our lives include the technologies that depend on scientific research. Our approaches to communicating, eating, becoming or staying healthy, moving from one place to another, reproducing, entertaining ourselves – all of these have been changed by the findings of scientific research. Perhaps less obvious, but not less important, are changes in how we think about ourselves and our world. The concepts and ways of thinking that scientists employ as they engage in research and seek to explain the phenomena they uncover tend to migrate out of the specialist settings in which they originate. Though they tend to be transformed in the process, these concepts and methods turn up in the broader culture as themes we draw upon to describe and explain the world.

Clearly, science is a big deal. This book is not just about science, though, it is about the philosophy of science. Even if you agree that one should care about science, you might wonder whether you should care about the philosophy of science. The point of this preface is to begin to persuade you that you should care about it. I will divide my arguments based on different types of readers. My first argument will be directed at those who either are or will become working scientists. My second argument will be directed at those people *and* at those who are interested in philosophy, but unsure about philosophy of science. My last argument will be directed at anyone, whether or not a scientist, whether or not philosophically inclined, who lives in a world as strongly influenced by scientific work as ours.

First, a caveat: These arguments are not meant to be fully persuasive as such. Full arguments regarding important practical matters – such as whether to devote one's valuable time to reading a particular book – can rarely be stated succinctly (and I do not wish to linger over-long here in the

preface, but to get on to the subject at hand). What I present here should rather be understood as the articulation of promissory notes that make more or less explicit the value of philosophy of science for assorted concerns that I take to be rather widely shared. It is my hope that reading through the rest of the book will leave you fully persuaded that this is a worthwhile subject matter. I grant that some measure of trust is required to get started.

Let me now address my arguments to those who are or will become working scientists. The argument here is quite simple: whether performing experiments, analyzing data, developing theories, building instruments, or writing papers, scientists at some point in their work implicitly *rely on* philosophical assumptions about their discipline and their subject matter. This is perhaps most easily shown in the case of those scientists who would most vocally *oppose* the importance of philosophy in science, who espouse a 'shut up and calculate' approach. To endorse such a view is to take a certain position regarding the value of attempting to interpret the meaning of a scientific theory; it is, in short, a philosophical claim to the effect that the value of theory resides only in its usefulness for calculating outcomes in application, whether experimental or technological. This position is a form of what philosophers call *instrumentalism*, and the instrumentalists constitute a long-standing and important tradition within philosophy of science. My aim is not to criticize instrumentalism, but only to point out that for the scientist there is no escaping philosophy. The question for the scientist is not whether to engage in philosophy, but only whether to do it well or badly.

One particularly important context in which philosophical considerations are of considerable importance to the practice of science concerns the use of statistics. Many experimental results are couched in terms that result from the use of statistical tools. Moreover, theoretical explanations of phenomena involving, for example, quantum mechanics, population biology, or meteorology invoke statistical concepts. These statistical notions are developed within frameworks that make certain assumptions about the meaning of such concepts as probability, randomness, and so on that are a matter of significant philosophical dispute. The use of statistics without an awareness of these philosophical issues is likely to result in confused or misleading statements about the meaning of one's research results or the character of natural phenomena. This book devotes significant attention to philosophical issues regarding the use of statistics.

Finally, and perhaps most fundamentally, I would like to advance the idea that the pursuit of scientific knowledge is already, at least in part, a kind of philosophical inquiry in the broad sense. This might look like a trick of definition-mongering, in which I have simply chosen to give 'philosophy' such a broad meaning that this claim becomes trivially true. But there is a good historical reason for taking empirical science to be continuous with philosophy in this way: much of what we would today call 'science' was once called 'natural philosophy.' It makes sense, too: among the aims of scientific theorizing are (arguably) the understanding of the processes of the natural world, the kinds of things or structures that might be found there, and how these would appear to us under various circumstances, react to our intervention in them, and develop over time. It does not seem to be a stretch to call what does these things a 'philosophy of nature.'

Perhaps, however, you are not now and are not planning to become a scientist. Let us suppose that you are instead someone with an interest in philosophy more generally. Any well-developed philosophical view of the world will include a philosophy of science at least implicitly. Suppose, for example, your philosophical interests center on whether human actions are truly chosen freely. Any position regarding this issue of free will must respond *somehow* to the results of scientific research in such domains as neurophysiology and cognitive science. To be sure, one possible response would be simply to ignore these scientific results and develop one's position on purely a priori grounds derived from introspection. The point is that whether one does this, or tries to be sensitive to the scientific findings in some more or less substantive manner, one must have *reasons* for treating the scientific results in the way one does, and these reasons must involve some views about the relationship between empirical science and philosophical theorizing – a philosophy of science, in short.

This brings us to "everyone else." Perhaps you have no particular scientific training and do not plan on receiving any. Perhaps you do not plan on engaging in extended philosophical reflection or argumentation as a significant part of your life's undertakings. For you, philosophy of science will remain important at least for practical reasons – as a kind of instrument of self-defense if nothing else. Politicians, marketers, even scientists themselves, will attempt to persuade you to act in certain ways based on what they present as 'sound science.' Without expertise in the relevant scientific field, you might find yourself wondering about the soundness of their claims. Although we

often have to rely on the testimony of experts to form our own opinions, and it is sometimes right to do so, the very question of when to thus defer to authorities and when to doubt them relies on the ability to critically evaluate such testimony. What is the nature of the authority of empirical science? What are the limits of that authority? What questions in general are important for scrutinizing a claimed scientific result? Training in philosophy of science should equip you with some basis for thinking soundly about these issues as they arise in the public discourse of and around science.

At its best, the scientific enterprise exemplifies the bold, yet responsible pursuit of knowledge. Appropriately weighted and clearly understood, the best of our scientific heritage should stand as a check on our too-often indulged propensity to let demagoguery and obfuscation lead us into unreasoned, even dangerous ways of thinking. An understanding of what makes for good scientific thinking may help us to hold ourselves to a higher standard in how we understand our world and our place in it.

Finally, studying science from a philosophical perspective should highlight and explicate a fact about scientific inquiry that, though in some sense obvious, might be easily forgotten when confronted with the difficulty of understanding the technical language in which unfamiliar scientific claims are presented: science is a human undertaking, an extension of our common search for understanding. Scientists extend our ordinary ways of knowing into realms that are remote in terms of distance, or size, or degree of abstraction, or complexity. Knowing just how they do so is important to evaluating the results of their inquiry. Moreover, perhaps understanding how inquiry is conducted in these remote realms will shed light back onto our common knowledge-seeking practices. In any case, to understand science philosophically is to understand something about what it is to be human.

Plan of the book

This book divides roughly into two parts. The first seven chapters ("Background and basic concepts") introduce both some important concepts that will be used throughout the discussion and some important figures and movements in twentieth-century philosophy of science who have helped shape the discipline and whose ideas remain relevant to current debates. We will consider the arguments of writers such as Pierre Duhem, Karl Popper, Rudolf Carnap, Thomas Kuhn, Imre Lakatos, and Paul Feyerabend. This will allow us

to gain some familiarity with the *falsificationist* (Popper) and *logical positivist* or *logical empiricist* (Carnap) schools of thought that loomed large in the first half of the twentieth century, while also considering later developments involving *post-positivist* thinkers (Kuhn, Lakatos, Feyerabend) that offered alternatives to the falsificationist and logical empiricist points of view. We will explore some of the issues and arguments that they pursued with the help of some examples from the history of science.

Chapters 8 through 12 ("Ongoing investigations") will survey some of the areas in which lively debates are ongoing in current philosophy of science. These surveys will seek to connect these current areas of debate with the historical precedents that are the focus of the first section of the book, but will emphasize foundational concepts and more or less recent developments. Chapters 8 and 9 will concern competing approaches to the use of probability ideas in science. Probability is crucial to the analysis of data and to the understanding of scientific reasoning. These chapters will, as you might expect, have the most mathematical content, but are meant to be self-contained. Chapters 10 and 11 will survey current debates concerning scientific realism vs. anti-realism, and explanation, respectively. Chapter 12 concerns the role of values and value judgments in science, a matter of long-standing concern that has recently seen a resurgence of interest.

Acknowledgments

I am very fortunate to have had the opportunity to write this book. I thought that I was simply creating a potentially useful teaching resource for myself and (hopefully) for others. The project turned into an opportunity to revisit long-standing interests and discover some new literature as well. I am grateful to Hilary Gaskin at Cambridge University Press for her sage guidance and immense patience. Her assistants Anna Lowe, and subsequently Rosemary Crawley, provided wonderful support. The final text benefited from the skillful copy-editing of Jon Billam.

As I have written, unwritten, and re-written, I have found many friends and colleagues willing to share their comments on various drafts. I would like to thank Carl Craver, Helen Longino, Deborah Mayo, Bill Rehg, Richard Richards, and Tom Pashby for their comments on assorted chapters of the book. Carl Craver invited me to guest lecture on Chapters 8 and 9 to a group of graduate students and postdocs at Washington University, and the ensuing discussion was very helpful. I am also thankful to the undergraduate students who have been enrolled in my Philosophy of Science survey class over the past couple of years; they have helped me field-test earlier drafts of the chapters in this book. The members of the History and Philosophy of Science Reading Group at Saint Louis University graciously read the entire text during the Fall of 2013, and our discussions were tremendously helpful. I am especially grateful to reading group stalwarts Scott Berman, Ben Conover, Mike Mazza, Jonathan Reibsamen, and Chong Yuan. In spite of having access to such collective wisdom, I probably have left in some things that are dubious or just plain dumb, and I have no one to blame but myself.

Dianne Brain has as ever advised, sustained, and encouraged me. I rely daily on her good judgment and patience. Our daughter Nora has provided many delightful distractions from my work on this text.

Abbreviations

Abbreviation	Meaning	Chapters
ACh	Acetylcholine	11
CDC	Centers for Disease Control	12
CERN	Conseil Européen pour la Recherche Nucléaire	9, 12
CL	Covering Law	11
CM	Causal-Mechanical	11
D-N	Deductive-Nomological	11
DHHS	Department of Health and Human Services	9
DNE1	Empirical requirement of Deductive-Nomological Account	11
DNL1-3	Logical requirements of Deductive-Nomological Account	11
DR	Differential Refrangibility	2
DR_{NH}	Differential Refrangibility restricted to the northern hemisphere	2
EPA	Environmental Protection Agency	9
GTR	General Theory of Relativity	9
HH	Hodgkin–Huxley	11
I-S	Inductive-Statistical	11
IBE	Inference to the Best Explanation	10
IID	Independent and Identically Distributed	9
LDD	Law of Doughnut Delectability	1, 2
LDD′	Alternate formulation of Law of Doughnut Delectability	2, 4
$LDD'_{a, b, c, d}$	The same, applied to individuals named a, b, c, d	2
LHC	Large Hadron Collider	9
MDC	Machamer, Darden, and Craver	11
MSRP	Methodology of Scientific Research Programs	6
NCM	Newtonian Classical Mechanics	6

NIH	National Institutes of Health	12
NMF	Naïve Methodological Falsificationism	6
NOA	Natural Ontological Attitude	10
NP	Neyman–Pearson	9
OD	Origin and Development derivation	11
PUDP	Principle of Uniformity of Doughnut Properties	1
PUN	Principle of Uniformity of Nature	1
RF	Resonant Frequency derivation	11
SCC	Squamous Cell Carcinoma	9
SMF	Sophisticated Methodological Falsificationism	6
SSR	*The Structure of Scientific Revolutions*	5
UD	Underdetermination	3

Part I

Background and Basic Concepts

1 Some problems of induction

1.1 Introduction

To the lay reader, much of what is written by scientists can seem barely comprehensible. Even to someone who has had some science courses in school, a sentence like "The M2 proton channel … has a 40-residue region that interacts with membranes consisting of a transmembrane helix (which mediates tetramerization, drug-binding, and channel activity), followed by a basic amphiphilic helix important for budding of the virus from cellular membranes and release (scission)" will seem as though it has been written in a language not quite identical to English (Fiorin, Carnevale, & DeGrado, 2010). As a result, the non-specialist may find much of what is accomplished by scientists mysterious and esoteric.

Philosophers of science have sometimes attempted to "humanize" scientific work by portraying scientific methods as extensions of our ordinary ways of knowing things. It is true that scientists use technologies both material and mathematical to make observations and draw conclusions that we could never achieve otherwise. Nonetheless, they observe, conjecture, infer, and decide just as we all do, if perhaps in a more systematic and sophisticated way. Although a certain kind of training is needed to understand some of the language scientists use to report their findings, those findings are not the result of magic or trickery, but of an especially refined and disciplined application of the cognitive resources we enjoy as a species.

But if scientists use and extend the cognitive abilities of ordinary knowers, they also inherit the philosophical problems of ordinary knowers. One of the thorniest and most discussed of these – the problem of induction – is the topic of this chapter. We will see how this general problem about knowledge arises in both non-scientific and scientific contexts. An example from the history of optics will show that scientific experimentation and careful use of

scientific methods can improve our ability to give well-reasoned answers to questions, but do not suffice to solve the problem of induction itself.

1.2 The problem of induction about doughnuts

The problem of induction is the problem of how we learn from experience. Consider how it arises in the life of a somewhat ordinary person, whom we shall call Zig. Zig likes doughnuts, let's suppose, and one day decides to try a new doughnut shop. This doughnut shop has introduced a new horseradish-flavored doughnut, something that Zig has never tried before. Zig expects, however, to like the horseradish-flavored doughnut. After all, he has enjoyed nearly all of the doughnuts he has eaten in his life (he has eaten many!). In fact, Zig believes there to be a general Law of Doughnut Delectability:

Law 1 (*LDD*) *All doughnuts are delicious.*

Since horseradish doughnuts are doughnuts, it follows that horseradish doughnuts must also be delicious.

Zig's reasoning here exhibits a logical trait of considerable significance. He is using an argument that is *deductively valid*, meaning that *it is impossible that the premises are true and the conclusion false*. This can be seen clearly enough in the case of Zig's argument. Suppose the *LDD* is true, and that it is true (as it seems it must be) that all horseradish doughnuts are doughnuts. Could there possibly be a non-delicious horseradish doughnut? Suppose there were. If we insist on the law that all doughnuts are delicious (first premise), then this non-delicious horseradish doughnut must not really be a doughnut after all. On the other hand, if we insist that all horseradish doughnuts really are doughnuts (second premise), then our non-delicious horseradish doughnut stands in contradiction to the law that all doughnuts are delicious. We simply cannot hang on to the truth of our premises while denying the conclusion of this argument.

Deductive arguments are sometimes said to be 'truth-preserving' because a deductively valid argument is guaranteed not to lead you from true premises to a false conclusion. This makes them powerful as logical tools of persuasion; as long as the person you are trying to convince agrees with you that your premises are true, she cannot resist the conclusion of a valid argument from those premises without contradicting herself. A deductively valid argument with true premises is said to be *sound*.

To his very great disappointment, Zig finds the horseradish doughnut revolting. He has just encountered a limitation of deductively valid arguments. Although the truth of their premises guarantees the truth of their conclusions, there is nothing about validity itself that guarantees the truth of the premises. Zig proceeds to make a new deductively valid inference: All horseradish doughnuts are doughnuts. Some horseradish doughnuts are not delicious (to say the least!). Therefore, some doughnuts are not delicious. With this argument, Zig has refuted the *LDD*.[1]

Now Zig begins to wonder: was that one horrendous-tasting horseradish doughnut a fluke? The next day he returns to the doughnut shop and orders another one. The second horseradish doughnut turns out to be equally vile. After ten more visits to the same shop, each time ordering a horseradish doughnut that he finds quite disgusting, Zig begins to visit other doughnut shops selling horseradish doughnuts. Zig proceeds to sample 12 horseradish doughnuts from each of 20 different doughnut shops and finds all 240 of them to be repellent.

Surely Zig is overdoing it! He has, you would think, more than enough evidence to justify an expectation that *any* horseradish doughnut he tries will not taste good to him. But suppose he remains skeptical of this claim. He freely admits the unpleasant taste of all horseradish doughnuts he has eaten in the past, but does not think that this has any bearing on what he should expect of those he has not yet tasted. What logical argument could we offer to persuade him?

The problem of induction amounts to just this skeptical problem: it seems that no amount of unpleasant experience with horrid horseradish doughnuts would suffice to justify any belief regarding the deliciousness or putridity of one that is yet to be tasted.

Consider the premises we have at our disposal for reasoning with Zig. “Look, Zig,” we might say. “You have eaten 240 horseradish doughnuts from a variety of doughnut shops in a variety of locations. You agree that each of those doughnuts was deeply unpleasant. Does it not follow that the next horseradish doughnut will be similarly awful?” Zig would be correct to reply “No, it does not follow. The premises you offer are consistent both with the

[1] You may have noticed that Zig could have saved the *LDD* if he had been willing to consider giving up the claim that all horseradish doughnuts are doughnuts. Maybe the terrible-tasting thing that he ate was not really a doughnut after all? We will return to the possibility of such evasions in Chapter 3.

next one being tasty and with it being terrible." In other words, the suggested inference to the conclusion that all horseradish doughnuts are not delicious (or even that just the next one lacks gustatory appeal) is not deductively valid. Our inference regarding the next horseradish doughnut is an inductive, not a deductive argument, and as presented it does not exhibit validity.

One might try to make it into a deductive argument by filling in some additional premises. Consider, for example, the following argument: All horseradish doughnuts in a large and varied sample are not delicious. All horseradish doughnuts that have not been tasted are similar to those in the sample with respect to their taste. Therefore, all horseradish doughnuts are not delicious.

Now we have a valid deductive argument, but we also have a *new premise*, and it is one that cannot be considered to report any observation that Zig has made. 'All horseradish doughnuts that have not been tasted are similar to those in the sample with respect to their taste' asserts a generalization about horseradish doughnuts, and Zig would rightly ask us to tell him what reasoning would justify his accepting this new premise. It is no more a deductive consequence of the uniformly bad taste of the horseradish doughnuts Zig has already eaten than is our original conclusion.

Perhaps this new premise can be supported by some separate argument that we think Zig should find compelling. One could, for example, regard it as a special case of a general Principle of Uniformity of Doughnut-Properties (PUDP): All doughnuts of a given kind resemble one another with regard to their taste. And this general principle, we could argue, is supported by past doughnut-eating experiences, not just of Zig, but of the general doughnut-eating public, which has found in the past that all raspberry-filled doughnuts tasted similar, all chocolate cake doughnuts tasted similar, all sour cream doughnuts tasted similar, etc.

Clearly, however, this justification for the PUDP is itself an inductive argument. Having declined to accept the conclusion of our original induction, Zig is under no logical compulsion to accept the conclusion of this argument either. We could try to patch up the logical gap with some other general principle such as the Uniformity of Food Properties, but then *this* principle will need to be supported by an inductive argument from uniformity in past food experiences.

At each stage in our effort to close the logical gap between premises and conclusion in an inductive argument, we will find ourselves invoking a

principle that in turn has to be justified inductively. Now either this series of justifications ends or it does not. If it ends, then it must end with us appealing to some general principle for which no further inductive justification is offered. This most general such principle is sometimes simply called the Principle of the Uniformity of Nature (PUN). Just how to formulate the PUN so that it is both strong enough to perform its logical task and plausibly true is difficult to say.[2] But we do not need to formulate it to see the problem: It is a general principle that cannot simply be true by definition (like 'carnivores eat meat') or be derived deductively from other premises that are true by definition. If it does not lend itself to such a deductive proof, and we have no inductive argument to support it, then it must simply be unjustified by any argument whatsoever, and so then are any of its consequences.

If, on the other hand, the series of justifications does not end, then we have an infinite regress of inductive arguments, so that the justification for our original conclusion can never be completed. Either way, the conclusion of any inductive argument (for there is nothing special about our doughnut example) turns out to lack any justification.

Note that in this illustration of the problem of induction we did not equip Zig with any specialized methods or concepts of the sort that scientists bring to bear on their research problems. The problem of induction appears not to be a problem of science so much as it is a problem of knowledge in general.

1.3 Reasoning about light: Isaac Newton's induction

Here, however, let us set aside the problems posed by our imaginary doughnut investigator's skepticism, and consider an actual bit of historically significant scientific reasoning. What we will see is that although the investigator – in this case, Isaac Newton – gave no explicit attention to the problem of induction in this general form that we have just considered, he devoted considerable attention to the question of just what he could justifiably infer from his observations, and the assumptions that would be

[2] The difficulties can be seen already emerging in the PUDP: How strongly do the doughnuts in a given kind resemble one another? If the principle asserts that they are exactly alike, then the PUDP seems to be false; if they are only somewhat similar in a very broad sense then the principle will be too weak to support the conclusions that are meant to be drawn from it.

required to justify those inferences. We might say that, although he did not concern himself with *the* problem of induction, he did concern himself with a number of specific problems of induction, i.e., problems connected with particular ways in which he might go wrong in drawing his conclusions.

> In a very dark Chamber, at a round Hole, about one third Part of an Inch broad, made in the Shut of a Window, I placed a Glass Prism, whereby the Beam of the Sun's Light, which came in at that Hole, might be refracted upwards toward the opposite Wall of the Chamber, and there form a colour'd Image of the Sun.

Thus begins Isaac Newton's "Proof by Experiments" of his second proposition in his 1704 work *Opticks*. The proposition in question states that "The Light of the Sun consists of Rays differently Refrangible" (Newton, 1979, 26). In other words, sunlight can be decomposed into rays of light that differ from one another in the angle at which they are refracted through a given medium.[3]

Newton proceeds to describe how he arranged his prism so that its axis was perpendicular to the ray of light incident upon it, and so that the angles of refraction at the incident face of the prism and at the face from which the rays exited the prism were equal (see Figure 1.1). He describes the image made by the refracted light on a sheet of white paper as "Oblong and not Oval, but terminated with two Rectilinear and Parallel Sides, and two Semicircular Ends" (Newton, 1979, 28), and proceeds to report in detail on its dimensions. He reports that when the refracting angle of the prism was reduced from its initial value of 64 degrees, the length of the image was reduced, but its breadth was unchanged. If the prism was turned so as to make the rays exiting the prism emerge at a more oblique angle, the image was lengthened. He notes some irregularities in the glass of his first prism, and thus tries the same experiment with different prisms "with the same Success." "And because it is easy to commit a Mistake in placing the Prism in its due Posture, I repeated

[3] Newton had first reported on his experiments with prisms in his inaugural lectures as the newly appointed Lucasian Professor of Mathematics at Cambridge in 1669 (Newton, 1984). His first published paper on the topic appeared in the *Philosophical Transactions of the Royal Society* in February 1671. There he describes the experiments as having been performed five years earlier "in the beginning of the year 1666," many years before writing the *Opticks* (Newton, 1978). The correct historical sequence of Newton's optical experiments and the trustworthiness of his narrative are matters of historical dispute (Guerlac, 1983; Newton, 1984).

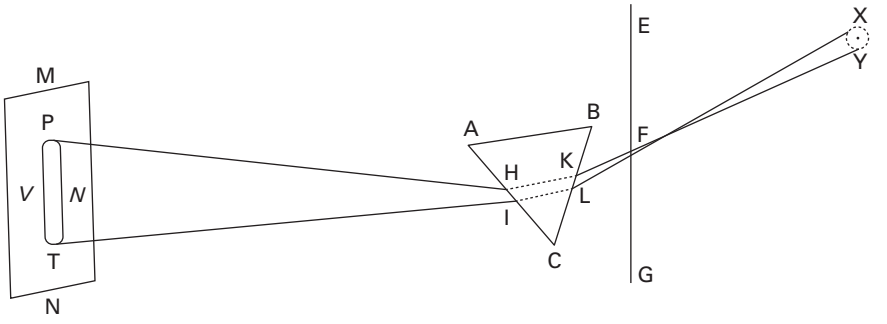


Figure 1.1 Newton's first experimental arrangement. *F* is the hole through which light from the sun (*XY*) passes, on its way to the prism (*ABC*). On a sheet of paper (*MN*) the light casts an image (*PT*) of the sun. (From Newton 1979, 27; by permission of Dover Publications.)

the Experiment four or five Times, and always found the Length of the Image" to correspond to that which he first reported (Newton, 1979, 30).

Newton then draws an intermediate conclusion from his measurements: The rays exiting the prism diverged from one another in such a manner as to constitute the length of the image, forming an angle "of more than two degrees and an half" (Newton, 1979, 31). That result Newton shows to be incompatible with "the Laws of Opticks vulgarly received," according to which the angles at which the rays exit the prism should diverge at an angle corresponding to the angle formed by the rays arriving at the hole from the extreme points of the sun's diameter: about half a degree. This would result in an image on the sheet of paper with a length equal to its breadth. What has gone wrong?

Newton points us to the answer by noting that the expectation of a round rather than elongated image would be correct were the rays that constitute the image "alike refrangible." "And therefore seeing by Experience it is found that the Image is not round, but about five times longer than broad, the Rays which going to the upper end ... of the Image suffer the greatest Refraction, must be more refrangible than those which to go the lower end" (Newton, 1979, 32).

There follows in the *Opticks* Newton's description of a series of experiments involving various arrangements of prisms through which Newton explores the phenomenon just described. Here I will just describe one such permutation that seems particularly relevant for understanding Newton's defense of his proposition that the sun's light consists of rays that differ in their

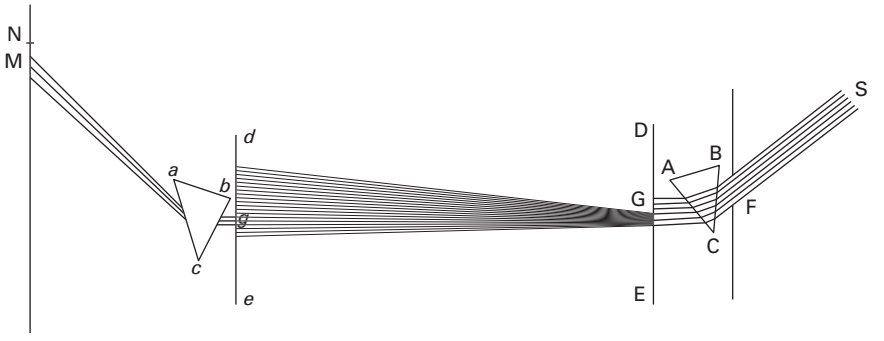


Figure 1.2 Newton's two-prism experimental arrangement. *S* is the Sun, light from which passes through an aperture *F* to prism 1 (*ABC*). A second aperture (*G*) can be used to select a portion of the refracted image of *S* to be sent through a third aperture (*g*) to prism 2 (*abc*). The refracted light from *abc* forms an image on the wall *MN*. (From Newton, 1979, 47; by permission of Dover Publications.)

refrangibility. In this arrangement (see Figure 1.2), Newton allowed sunlight to enter through "a much broader hole" before passing into a first prism, the orientation of which about its axis could be adjusted. The refracted light is then directed at a board with a hole about one-third of an inch in diameter. The beam of light thus formed is then projected onto a second board "and there paint[s] such an oblong coloured Image of the Sun" as had been found in the previously described experiment. This second board also has a small hole in it, and light that passes through that hole is then sent into a second prism, with its axis parallel to that of the first prism, which is held at a fixed position.

By varying the orientation of the first prism, Newton effectively sends different parts of the spectrum created by the first prism into the second prism. In this way he is able to more stringently focus on the question of whether these rays that emerge at different angles from the prism really do, *considered separately from one another*, refract at different angles. To test whether this is so, he observes the location on the opposite wall of the room at which the resultant ray of light arrives, having passed through both prisms. He found that indeed light from the higher part of the spectrum projected onto the second board refracted through the second prism to a higher location than light from the lower part of the spectrum, while light from intermediate parts of the spectrum arrived at intermediate locations on the wall. Newton thus is able to show that, although each ray arrives at the second prism at the same

angle, "yet in that common Incidence some of the Rays were more refracted, and others less. And those were more refracted in this Prism, which by a greater Refraction in the first Prism were more turned out of the way, and therefore for their Constancy of being more refracted are deservedly called more refrangible" (Newton, 1979, 48).

Clearly, the proposition that Newton defends is a general one. He does not claim merely that the particular rays of solar light that he observed are composed of rays that vary in their refrangibility. Neither is his claim limited to the sun's light as observed in Cambridge, or as observed in the late seventeenth century. Indeed, although his explicit claim makes reference to the light of our Sun, his motivation for this research and subsequent application of his findings has to do with limitations on the resolution of telescopes, which he considers to arise at least in part from the differential refraction of the component rays of light from celestial objects, including stars. Clearly, then, he thought that the same proposition applied to light coming from at least some stars other than our own sun. His proposition states much more than can be summarized by simply noting what he observed in the particular instances he describes.

1.3.1 Newton's problems of induction

It is striking, therefore, to consider how little Newton was concerned about the problem of induction as we articulated it above. There is no evidence that Newton was at all worried about whether light might behave differently on Mondays than on Thursdays, or that his results might come out differently if he tried his experiment in the country instead of in town, or in Bristol rather than Cambridge. If anyone else worried about these matters, Newton's argument includes no overt reasoning that would allay those concerns. Given just the description of the experiments that Newton did perform, we would have to say that it remains *logically possible* that the Sun's light observed at different times or under different conditions does *not* consist of rays that differ in their refrangibility.

But Newton clearly is concerned about other possible ways in which his claim might go wrong, and designs his experiments explicitly to rule out these sources of error. For example, Newton considers the oblong (rather than circular) shape of the image created by the prism to be significant, as it differs from the prediction from "the Laws of Opticks vulgarly received" (Newton,

1979, 31). What if that shape were somehow an artifact of the particular prism that Newton used, perhaps because of a defect in the glass? Newton must have been concerned about *this* possible error, because he describes how he repeated the experiment using different prisms and got the same result. What if Newton had made a mistake in lining up his prisms in the correct configuration? Again, he repeated his experiments to rule out small but unnoticed errors in his alignment of the prisms.

These two possible sources of error – flaws in the prisms or errors of prism placement – might be thought of as *local* in the sense that they concern details of the particular instances from which Newton seeks to draw conclusions. But sources of error that are more *global* in nature are also relevant for him.

For example, why does Newton bother with the second experimental arrangement just described, in which he passes light refracted from the first prism through two more apertures before it enters the second prism? Although Newton does not state explicitly his motivation for this experiment, one might consider it as addressing a concern about a potentially systematic misunderstanding of the phenomenon thus far studied. Although the image resulting from refraction by the prism clearly indicates that light exits the prism at a range of angles, one might worry that somehow the phenomenon cannot really be analyzed in terms of the differing properties of refrangibility of rays in the incident light. Perhaps the spread somehow involves all of the light together, such that if some part of the spectrum is taken away, the whole phenomenon is destroyed. A worry of this sort would be addressed by Newton's experimental arrangement, which allows him to demonstrate that individual segments of the spectrum consist of light that, just by itself, refracts at an angle different from that of the light in other parts of the spectrum.

Another sort of worry might arise from concerns about just what is implied by Newton's conclusion. Just what does Newton mean by saying that the Sun's light "consists of Rays differently Refrangible"? Here one might take him to be making a claim about the nature of light itself. In particular, one might worry that the term "rays" points toward a view of light as a kind of particle, the trajectory of which is described by a ray. But if Newton means to imply such a view of light, then he would seem to be going well beyond what his experiments can show, since there does not seem to be anything in these experiments as described that would rule out other theories about the nature

of light, such as that it consists of some kind of wave (theories of this sort had been advocated by Newton's contemporaries Christian Huygens and Robert Hooke).

It was in fact Hooke himself who challenged Newton on this score, charging that his conclusions about color were part of a particle theory of light. In response to Hooke, Newton drew a distinction between hypothetical theories about the nature of light that would causally explain what he had observed experimentally, and a description of those properties of light that he considered to be revealed in his experiments. Recognizing that his observations could be explained by "Mechanicall" theories other than the particle account he himself favored, he wrote

> therefore I chose to decline them all, & speake of light in generall termes, considering it abstractedly as something or other propogated every way in streight line from luminous bodies, without determining what that thing is. (quoted in Shapiro, 2002, 232)

Thus Newton denies that we should draw any inferences from his use of the term "Rays" to conclusions about the nature of light itself. He thus secures his argument against objections against any such theories about light by making his conclusion independent of such assumptions.

To solve *the* problem of induction, Newton would have had to either show how, through an inductive argument, he could eliminate all possibilities of error in his conclusion, or else to show, without falling into a regress or a vicious circle, how his conclusion could be justified *without* eliminating all possibilities of error. Newton did not attempt either of these things, and thus did not attempt to solve the problem of induction. He did, however, attempt to solve problems relating to *particular* ways he might go wrong in drawing his conclusion. The example of Newton's optical experimentation illustrates how, even if one does not worry overtly about the problem of induction in general, a convincing experimental argument for a general claim does require that one be able to rule out sources of error that are judged to be relevant to one's conclusion.

1.4 Conclusion

In philosophical discussions of the problem of induction, attention typically focuses on cases like our doughnut example, in which the inductive inference

consists of a simple generalization from observations made in some finite sample to some broader population of instances (sometimes called *enumerative induction*). We have seen how Newton, although he does generalize in his conclusion, undertook to justify his conclusion by using experimental *probing* to explore the various ways in which his conclusion might be in error. This *active* error-probing[4] reflects more accurately the reasoning of experimental science than do philosophy-textbook stories of passive extrapolation like Zig and the doughnuts.

This still leaves us with some tough questions: Which possible sources of error should one consider relevant? Perhaps Newton justifiably disregards worries that the light from the Sun might behave differently at different times or in different places, but why disregard those possibilities but explicitly address those about the possibilities of error in the alignment of prisms or of flaws in the prism glass? If one argues that Newton justifiably assumes that the Sun's light behaves uniformly in different places and at different times, then one has to give a justification for this that does not thrust us back into a regress or vicious circle of inductive arguments. On the other hand, if the difference between these possible errors is simply that no one, neither Newton nor his audience, was worried about the former but that they were – or plausibly would be – worried about the latter, then Newton's justification seems to depend on nothing other than the opinions of his contemporaries. That would have the rather unpalatable consequence that one's inductive arguments would be stronger, all else being equal, simply in virtue of one's audience being less critical!

Worse, it seems that although Newton does not worry about the problem of induction, he *should*! Simply ignoring a problem does not make it go away, and we are still left with the unsatisfactory situation of relying on an inductive form of argument that we ultimately seem unable to defend or justify. In our next chapter we will consider an approach to science that purports to do without induction entirely, but allows science to proceed without making use of a method that we cannot justify.

[4] For more on this, see Chapter 9, Section 9.4.

2 Falsificationism: science without induction?

2.1 Growth through falsification

The philosopher of science Karl Popper considered the problem of induction so grave as to require the *elimination* of induction from the methods employed in science. He went so far as to assert that "in my view there is no such thing as induction" (Popper, 1992, 40).

In light of our discussion of inductive arguments in the previous chapter, such a proposal may sound outrageous. Surely, it might seem, we *need* inductive reasoning in science; it is by such reasoning that we confirm or support general propositions such as natural laws. That is how we learn from experience, fostering the growth of scientific knowledge.

Scientific knowledge does indeed grow, according to Popper. Understanding this growth is for him the central problem of the philosophy of science and of epistemology (the theory of knowledge) more generally. But he regarded the characterization of this growth in terms of induction as deeply mistaken. The belief that inductive arguments from the results of particular observations serve to confirm, support, or justify general empirical propositions – a belief we can label 'inductivism' – is a dangerous illusion. That inductivism is the wrong way to understand this growth can be seen most clearly, according to Popper, if we consider an alternative view, which he calls *falsificationism*.

The key to falsification lies in a logical asymmetry that we noted previously. Although the bulk of our story about Zig and the doughnuts was concerned with the relationship between Zig's many unpleasant observations of particular horseradish doughnuts and the proposition that all horseradish doughnuts are distasteful, the story began with an apparently quick and successful *refutation* of a proposition that Zig initially believed: All doughnuts are

delicious (the Law of Doughnut Delectability). It took only a single foul-tasting horseradish doughnut to reveal the falsity of the *LDD*.

2.1.1 The logical structure of generalizations

Such a falsification can even be represented in terms of a valid *deductive* argument. It will help us to see how such an argument works if we translate the *LDD* into a more precise logical form.

The logical form of *LDD* can be specified if we think of it as a generalization of a *conditional*. A conditional takes the form 'If p, then q', where p and q should be understood as *variables* that stand in place of propositions that could be expressed by sentences all by themselves, like 'the snowman is melting.' The part (p) of the conditional that immediately follows 'if' is called the *antecedent* and the part that follows 'then' (q) is called the *consequent*. *LDD* can be thought of as making completely general the conditional 'if it is a doughnut, then it is delicious.'

There is something a bit odd about statements like 'it is a doughnut,' 'it is delicious,' or 'if it is a doughnut, then it is delicious.' Absent any context to determine what it is that 'it' refers to, such statements lack any definite truth values. Whether the statement 'it is a doughnut' is true depends on what 'it' refers to when that statement is made.[1] To turn the conditional 'if it is a doughnut, then it is delicious' into a universal generalization, we preface it with a universal 'for any' quantifier:

Law 2 (*LDD′*) *For any thing whatsoever, if it is a doughnut, then it is delicious.*

LDD′ states exactly the same thing as *LDD*, but in the form *LDD′*, we can see more clearly that the law makes a statement not only about doughnuts, but about literally everything. What it states is this: Select, however you like,

[1] Logic therefore does not consider 'if it is a doughnut, then it is delicious' to be a sentence, and the same applies to its antecedent and consequent. In spoken language, this expression *could* be uttered so as to give it the meaning of its generalized form *LDD* (consider what is probably meant by someone who says something like, "Hey, if it has cheese on it, then I like it"). Logicians attempt to tame these idiosyncrasies of natural languages by putting everything into symbolic forms like '$\forall(x)$[Doughnut(x) $\supset$ Delicious(x)]', thus separating in symbols the form of the conditional ('Doughnut(x) $\supset$ Delicious(x)', which is read as 'if x is a doughnut, then x is delicious') from the operation of generalization ('$\forall(x)$', which is read as 'for every thing x').

Table 2.1 *Doughnuts and other things.*

name	doughnut?	delicious?
a	yes (chocolate)	yes
b	yes (horseradish)	no
c	no (asparagus)	yes
d	no (submarine)	no

anything from any location in the universe. Is the thing selected a doughnut? If so (whether it is a chocolate, blueberry, or horseradish doughnut), then it is delicious. If not (whether it is an electron, a lemon, or Luxembourg), then it may be delicious or it may not.

To understand the behavior of logical conditionals like 'if it is a doughnut, then it is delicious' consider how we might apply *LDD′* to individuals. Suppose we obtain four things, by whatever means. We can give these four things arbitrary names, so that we can talk about them without committing ourselves to any assumptions about what kinds of things they are. (Imagine, if you like, that you are presented with four boxes of arbitrary size and are simply told that each contains one thing to which *LDD′* will now be applied.) Let's call these things a, b, c, and d. Applying *LDD′* to them yields the following conditional sentences:

LDD'_{a} If a is a doughnut, then a is delicious.
LDD'_{b} If b is a doughnut, then b is delicious.
LDD'_{c} If c is a doughnut, then c is delicious.
LDD'_{d} If d is a doughnut, then d is delicious.

Conditionals like these are false when their antecedent is true and their consequent is false. Otherwise, they are true. Suppose that we open our four boxes and inspect their contents. We want to know of each thing whether it is a doughnut and whether it is delicious. Our data are given in Table 2.1.

Given these data, LDD'_{b} is false, and the other three conditionals are true.[2]

[2] It might seem odd to say that it is true that if d is a doughnut, then d is delicious, given that the name d refers to a submarine. Classical logic assumes, however, that all well-formed sentences, however peculiar their content, are either true or false, and treating LDD'_{d} where d is a submarine as *false* would lead to the unfortunate consequence that one could falsify a generalization about doughnuts by making observations of submarines.

2.1.2 Falsification revisited: a *modus tollens* argument

Now consider the import of our observation of b for the sentence *LDD′*. From the observation that b is a doughnut and b is not delicious, it follows that there is a thing such that it is a doughnut and it is not delicious. But it is a logical consequence of *LDD′* that it is not the case that there is a thing such that it is a doughnut and is not delicious. We can therefore express the refutation of *LDD′* in the following argument:

(1) If *LDD′* is true, then it is not the case that there is a thing such that it is a doughnut and it is not delicious.
(2) There is a thing such that it is a doughnut and it is not delicious.
(3) Therefore, *LDD′* is not true.

In this argument premise (1) must be true because the consequent is a logical consequence of the antecedent. Premise (2) is taken to be true on the basis of the observation of a particular horseradish doughnut (more about this step below). These two premises suffice to show, using a deductively valid argument, that (3) *LDD′* is false.

One may thus *falsify* a general empirical proposition, according to Popper, without appeal to induction. For such falsification one can rely on a deductive form of argument known as *modus tollens*. Such arguments take the form *If p then q. It is not the case that q. Therefore, it is not the case that p.*

2.1.3 Newton's experiment as falsification

It might seem that Popper's rejection of inductive arguments must lead to the rather odd conclusion that Newton's achievements in his optical experiments – an icon of the 'scientific revolution' – were an illusion. To draw this conclusion would, however, be to misunderstand the falsificationist view. Newton did indeed advance scientific knowledge with his prism experiments, but not by inductively confirming his claim about light. In Popper's view, Newton's contribution to the growth of knowledge has two facets: a *success* and a *failure*.

First, Newton *succeeded in falsifying* a claim about light. Recall how he noted that the "Laws of Opticks vulgarly received" lead to the expectation in his first experiment that a round image will be formed on the screen by the refracted light from the Sun. More precisely, from the hypothesis that all light

is refracted by the same amount for a given refracting medium and a given angle of incidence, Newton derives the prediction that light refracted by the prism from the Sun will form a round image on the screen. The oblong image that Newton found when he performed his first experiment conflicts with this prediction, leading, via *modus tollens*, to the denial of the hypothesis that all light is refracted by the same amount for a given refracting medium and a given angle of incidence. So Newton succeeded in establishing the denial of a proposition that many had previously thought true.

Second, Newton *failed to falsify* his own hypothesis (*DR*, for Differential Refrangibility) about light: that light from the Sun consists of rays that differ in their refrangibility. According to the falsificationist, such a failure contributes to the growth of scientific knowledge not simply because a falsification did not happen (which would also be true if the experiment had never been performed, or if it were designed so poorly as to be irrelevant to the hypothesis in question), but because Newton's experiments were of a sort that might have falsified this hypothesis, yet they did not. In other words, Newton's experiments attempted, but failed, to falsify the hypothesis of differential refrangibility.

Fair enough, you might be thinking, but Newton's experiment also failed to falsify other hypotheses about light's refrangibility, such as the hypothesis (DR_{NH}) that light from the Sun that arrives in the northern hemisphere consists of rays that are differentially refrangible. In the problem of induction, we saw that any finite set of individual observations was logically compatible with a multitude of different generalizations, and one might think that a similar problem is surfacing here.

2.1.4 Corroboration

Popper's response to this issue is crucial to his account of the growth of knowledge. To characterize such growth, we need to attend not only to those hypotheses that have survived potential falsification, but to the *corroboration* of hypotheses. Corroboration is a technical term introduced by Popper that is meant to be free of any of the inductive valences of terms such as confirmation, support, or verification. Popper stipulates that the degree of corroboration of a hypothesis is a function of both the extent to which the hypothesis has survived potentially falsifying tests, and of the 'boldness,' or more precisely, *degree of falsifiability* of the hypothesis itself.

One theory is more falsifiable than another if, roughly speaking, it affords more opportunities for being falsified. Both *DR* and DR_{NH} could have been falsified by Newton's particular experiment, but *DR*, because it asserts something about light from the sun everywhere, affords many *more* opportunities for falsification. Experiments similar to Newton's could be carried out in the Southern Hemisphere, on the Space Station, or on other planets. All of these locations would provide opportunities to attempt a falsification of *DR*, but such experiments would be irrelevant to DR_{NH}. By contrast, any experiment that tested DR_{NH} would necessarily also test *DR*. Thus, while it is true that both *DR* and DR_{NH} survived falsification in Newton's experiment, they are not equally corroborated by that experiment.

Theories that make broader, more general claims, and theories that are simpler in their description of the world will tend to be more falsifiable because they have larger classes of potential falsifiers. A more falsifiable theory "*says more* about the world of experience" than one that is less falsifiable because it rules out more possible experimental outcomes (Popper, 1992, 113; emphasis in original). Popper's analysis of falsifiability centers not on what a hypothesis says *will* happen, but on what it forbids, on the experimental results that, according to that theory, will never be produced. The more potential observations a theory rules out in this way, the more falsifiable it is.

Thus Popper endorses a methodology that advises us to prefer always those theories that are the most highly corroborated – i.e., those that are the most falsifiable ones to have survived testing so far.

2.2 Falsificationism, demarcation, and conventionalism

2.2.1 Demarcation and the aim of science

By his own account, Popper began to formulate his philosophy of science in Vienna around 1919, when "the air was full of revolutionary slogans and ideas, and new and often wild theories" (Popper, 1965, 34). He noted a contrast amongst some of these "wild" theories. Einstein's General Relativity completely reconceptualized the phenomenon of gravity. The first experimental test of General Relativity was performed in 1919, a test that Popper regarded as precisely the kind of potential falsification that made for good science. By contrast, other widely discussed theories – for example the Marxist theory of

history and the psychoanalytic theory of Sigmund Freud – seemed to be so flexible as to be able to explain any potentially relevant observation whatsoever, even though the individual case might seem initially problematic.

In the case of Marxist history, Popper wrote that, although the theory in its early formulations made testable predictions, when these predictions were in fact falsified, the advocates did not respond by giving up the theory but instead "re-interpreted both the theory and the evidence in order to make them agree" (Popper, 1965, 37). He considered Freudian psychoanalysis not to be susceptible of falsification at all: "There was no conceivable human behaviour which could contradict [it]" (Popper, 1965, 37).[3]

Popper came to see the challenge posed by this contrast between Einstein's and Freud's theories in terms of what he called the 'demarcation problem': how to distinguish science from pseudoscience (or pre-science). He postulated falsifiability as a criterion for making precisely this demarcation: "the criterion of the scientific status of a theory is its falsifiability, or refutability, or testability" (Popper, 1965, 37).

By stipulating this criterion, Popper's philosophy of science makes a claim not only about what makes a good theory, but also about what science is *about* or, more precisely, what constitutes the *aim of science*. For Popper, to engage in science is to pursue knowledge by *considering only falsifiable theories* (i.e., theories that entail sentences denying the occurrence of certain possible experiences) and by then proceeding to *attempt the falsification of those same theories* (by producing the circumstances under which those experiences might be produced). Popper sought in this way to articulate an ideal that would stand in opposition to the possibility that scientific theories might become matters of dogma, in which adherence to a theory is adopted simply as a matter of decision not subject to criticism. Popper referred to the latter danger, exemplified for him in the attitudes of some adherents to the Marxist theory of history and Freudian psychoanalytic theory, as a form of *conventionalism*. As great as Einstein's General Theory of Relativity is, according to Popper, it would be a breach of good scientific conduct to accept that theory as a matter

[3] Popper did not conclude from this that Freud's ideas had no value. He allowed that, although Freud's theory is untestable, it "is of considerable importance, and may well play its part one day in a psychological science which is testable" (Popper, 1965, 37). He compared such "pre-scientific" ideas to myths that could provide suggestions to later scientists.

of convention, simply because of a prior decision to do so. To do so would prevent the growth of knowledge.

It might therefore seem that Popper's philosophy of science rests on a certain normative claim (i.e., a claim about what one *ought to* do or believe): that it is *better* to engage in scientific inquiry by attempting to falsify theories than by attempting to confirm them through inductive arguments or by attempting to defend them against refutation. This yields a potential problem for the falsificationist, who cannot accept this proposition (which is a generalization) on the basis of an inductive argument. Yet it also seems unfalsifiable, for what possible observation would falsify it?

2.2.2 Rules as conventions

The first thing to be said in response to such an objection is that Popper denies that falsificationism is based on a proposition that one must believe. Instead, falsificationism is constituted by a commitment to a certain set of rules. These are rules for doing science. The rules that govern any kind of activity are, according to Popper, conventions. A proposed set of rules for doing science can no more be established by logical arguments than can a proposed set of rules for playing chess. Popper chooses falsificationist conventions for their effectiveness in preventing scientific theories themselves from becoming conventions.

In other words, if science is to be a rule-governed activity, then one *must* be a conventionalist about methodology. Given that conventionalism about method is unavoidable, Popper advocates choosing those conventions for the conduct of science that preclude conventionalism about the *content* of science.

A second important point is that, although Popper acknowledges that falsificationism is not falsifiable in the same sense that Einstein's General Theory of Relativity is falsifiable, it is susceptible to *criticism*. Popper offers susceptibility to criticism as a criterion more appropriate to the evaluation of conventional rules and philosophical positions more generally. In particular, one can ask of a proposed methodology for science whether the pursuit of science in keeping with its rules leads to desirable results. Popper writes that

> My only reason for proposing my criterion of demarcation is that it is fruitful: that a great many points can be clarified and explained with its help. . . . It is only from the consequences of my definition of empirical science, and from

> the methodological decisions which depend upon this definition, that the scientist will be able to see how far it conforms to his intuitive idea of the goal of his endeavours. (Popper, 1992, 55)

We can thus see that Popper embraces (because he believes he must) a certain kind of conventionalism regarding science: namely, that scientific inquiry is directed at enabling the pursuit of *modus tollens* arguments such as that considered above to be brought to bear on highly falsifiable hypotheses. Let us now look a little more carefully at the premises of such arguments.

2.3 The empirical basis

2.3.1 Avoidable falsifications

You may recall that Zig was very quick to accept the falsification of *LDD* after tasting a single non-delicious horseradish doughnut, but very reluctant to accept the generalization that all horseradish doughnuts are distasteful. Suppose he had been inclined oppositely, however, and was loathe to give up his belief in the *LDD*. Does logic really force him to give it up? Of course, he has to admit that *LDD* really is a generalization, and so it applies to the object that he just tasted. But this falsification also depends on two further claims about that object: that it (a) *is a doughnut* and (b) *is not delicious*. Accepting these claims enables him to conclude that there is something that is both a doughnut and not delicious, in direct contradiction to the logical consequence of *LDD′* that there is no such thing. Although we might think it strange to dispute whether something really is a doughnut or not, Zig could dig in his heels on this point. Although it is a doughnut-shaped object sold in a doughnut shop, perhaps it is not really a doughnut. Perhaps it is a bagel disguised as a doughnut. If we show him the dough being made and then fried, rather than baked, he might insist that we did not use the kind of flour required for a doughnut (perhaps if it were made with rice flour, it would not be a proper doughnut). When all else fails, he can resort to conspiracy theories or plead hallucination. Perhaps none of these responses would seem reasonable to us, but the point is that they could be made in a manner that is logically consistent with continuing to believe that the *LDD* is true.

The consequence of this is that, although Popper jettisoned induction in favor of falsification as the engine of the growth of knowledge at least in part because he considered induction logically suspect, it remains the

case that falsification is not achieved by logic alone. The decision to accept the premises of a falsifying *modus tollens* argument is not forced by logic.

2.3.2 Deciding to accept a falsifying statement

Popper was forthright on this point. In order for scientific inquiry to proceed by enabling the pursuit of falsifications, scientists must *accept* certain statements as potentially falsifying premises, even though those premises are not forced upon them by logic. These accepted statements form what Popper calls *the empirical basis* (Popper, 1992, 93–111). Popper requires statements comprising the empirical basis to take the form of singular existential claims (such as 'there is a doughnut that is not delicious in Minot, North Dakota' or 'there is a refracted image of the Sun that is not round in Cambridge, England').

Now one might suppose that, even though such a statement is not forced upon us by logic, it is nonetheless *justified by experience*. Newton did, after all, *observe* that the image of the Sun cast upon the screen was not round, and surely that experience of his constitutes support for the claim that there was, in his experimental setup, an image of the Sun that did not display the shape predicted by the theory of uniform refrangibility of light. Popper, however, considers this view a mistake. Statements, he insists, are not supported by anything except other statements that logically entail them. An experience is a psychological phenomenon, not a statement. We therefore cannot even *apply* the rules of logic to experiences. Justification and support are the wrong categories for talking about the relationship between experiences and the statements we accept as describing those experiences.

Accepting a statement, according to Popper, should be regarded as a *decision* made in response to an experience. As responses, we might consider some decisions more appropriate than others in the light of a given experience, but these are practical, not logical judgments. The decision to accept a statement as part of the empirical basis in particular should be judged as either fruitful or not, and in this sense, the empirical basis has a conventional status, like rules of method. Decisions about the empirical basis have to be judged with regard to how well they serve the aim of enabling the pursuit of falsifying arguments as the engine of the growth of scientific knowledge. Popper's comments on this are worth quoting in full:

> The empirical basis of objective science has nothing 'absolute' about it. Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being. (Popper, 1992, 111)

2.4 Conclusion

Popper's falsificationist ideas remain very influential. The criterion that a scientific theory must be falsifiable receives wide support among scientists, although they may not be aware of the full philosophical framework of falsificationism that Popper built around that criterion. Popper's attempt to characterize the growth of scientific knowledge in a manner that makes no appeal to induction has also confronted a number of challenges. For example, Wesley Salmon has argued that falsificationism cannot make sense of the apparently rational *use* of highly corroborated theories for the purpose of making practical predictions (Salmon, 1981).[4] In the next chapter we will consider a challenge that confronts both inductivists and falsificationists and that will enter into most of the philosophical debates that lie ahead: the problem of underdetermination.

[4] An example is the planning of trajectories of interplanetary unmanned spacecraft, such as the Cassini-Huygens mission. These probes are able to travel to distant planets such as Saturn by using a 'gravitational slingshot' or 'gravity assist,' the execution of which demands the use of a theory of gravity, such as Newton's or Einstein's.

3 Underdetermination

3.1 Introduction

We might think of the problem of induction in terms of a logical *gap* between the premises of an inductive argument and its conclusion. So that scientific knowledge could be achieved without having to leap such gaps, Karl Popper proposed a theory of scientific method that focused on falsification rather than induction. In doing so, however, he had to confront the fact that, although falsifications can be *represented* with a gapless, deductively valid argument, logic does not require one to make the inference thus represented. Another gap, however, looms over both inductivism and falsificationism.

We saw in the previous chapter how Zig could have avoided falsifying the Law of Doughnut Delectability by insisting that the non-delicious object he tasted was not, after all, really a doughnut. In such a case the logical gap exists because logic does not force on the observer a particular description of one's observation. Another logical gap arises because the logical connection between the hypothesis or theory being tested and the description of the data or observations is not direct. One needs additional premises (often called *auxiliary hypotheses* or *auxiliary assumptions*) in order to establish a logical connection between a theory and a description of a possible observation. Falsification starts with a determination of the falsehood of the observation that is expected assuming that *both* the hypothesis under test and the auxiliary assumptions are true. The resulting *modus tollens* argument therefore only shows that *it is not the case that both the hypothesis under test and the auxiliary assumptions are true*. But this is compatible with the hypothesis under test remaining true while one or more of the auxiliary assumptions is false.

The existence of such a logical gap is often expressed as the *underdetermination thesis*: Theories are underdetermined by data. The underdetermination thesis is often associated with the French physicist and historian/philosopher

of science Pierre Duhem, who gave a prominent articulation of it in his book *The Aim and Structure of Physical Theory*, first published in 1906 (Duhem, 1991). In this chapter we will consider Duhem's underdetermination thesis, look at how his thesis was subsequently taken up and modified, and finally consider its bearing on our understanding of scientific knowledge. In later chapters we will discuss how certain approaches to scientific inference attempt to confront the problem posed by the underdetermination thesis.

3.2 Duhem's thesis

Contemporary philosophers of science often associate the underdetermination thesis with Pierre Duhem, but some of the distinctive features of Duhem's own thesis are often forgotten. Before we turn to the contemporary discussion of underdetermination, let us consider Duhem's original argument. First we will look at an example that Duhem uses to illustrate underdetermination. Then we will note two significant features of Duhem's discussion: (1) he regards the thesis as having special importance for physics, as opposed to other sciences; (2) he takes the thesis to express a *problem* confronting the physical investigator, and then discusses in a general way the resources needed for the investigator to *solve* that problem.

We might express Duhem's thesis as follows:

Thesis 1 (UD) *Given that a set of premises* $\{P_1, P_2, \ldots, P_n\}$ *deductively entails a statement O describing a possible observation, and that from experimental observation one concludes that the statement O′ is true, where O′ entails that O is false, it follows as a deductive consequence that the conjunction '*P_1 *and* P_2 *and … and* P_n*' is false. It does not follow from this that any particular member of* $\{P_1, P_2, \ldots, P_n\}$*, including the hypothesis that one seeks to test, is false.*

This thesis follows in a straightforward way from basic rules of deductive inference. Thus, the real issue we wish to take up is not whether UD is true (which it certainly is), but what import it has for the conduct of scientific inquiry and our understanding of its results.

First, however, we should pause to note that, as articulated here, UD is really a thesis that applies in contexts in which a theory is being evaluated by *testing its consequences*. It will therefore be relevant to any methodology of science, inductivist or falsificationist, in which such testing plays a role.

3.2.1 A test of Newton's emission theory of light

To understand the argument that Duhem makes, let us consider Duhem's own example involving another aspect of Newton's theorizing about light. You may recall from Chapter 1 that in defending his conclusions about the differential refrangibility of light against Hooke's objections, Newton argued that his conclusions did not depend on any substantive assumptions about the nature of light itself. Nonetheless, Newton did have a theory about the constitution of light. According to his 'emission theory,' light consists of very small projectiles emitted by bodies such as the Sun at very high velocities. These projectiles penetrate through transparent bodies, which in turn reflect and refract the light as a result of the action of the particles that make up the bodies on the light particles. In effect, what Newton proposed was that the matter of a transparent medium exerts an attractive force upon the light, so that light passing into a more dense medium would speed up due to the greater attractive forces exerted. On this account, when light passes from one transparent medium into another (say, from air to water), the index of refraction (from which one can calculate the angle of refraction) is equal to the velocity of light in the second medium divided by the velocity of light in the first. That light should travel faster in water than in air became an object of experimental interest when the French physicist Françoise Arago proposed a method to test this consequence of Newton's theory.[1]

The experiment itself was not performed by Arago, but separate measurements by Hippolyte Fizeau and Léon Foucault led to the conclusion that, contrary to the prediction from Newton's theory, light travels more slowly in water than in air.

Granting that others originated the basic idea behind it, Foucault's experiment is noteworthy for its technical precision and ingenuity of design (see Figure 3.1).[2] Sunlight entered the apparatus through a rectangular aperture containing a single vertical wire. Light then passed through an achromatic lens L on its way to a plane mirror m. The mirror m rotates (with the help of a small steam engine) around a vertical axis passing through its center. When m is appropriately positioned, light reflects from m to either of two concave

[1] Arago's proposal had its source in a similar idea due to the English physicist Charles Wheatstone, who had used a rotating mirror to measure the velocity of electric current.

[2] See (Preston, 1890; Tobin, 2003), on which the following description is largely based, for a helpful discussion.

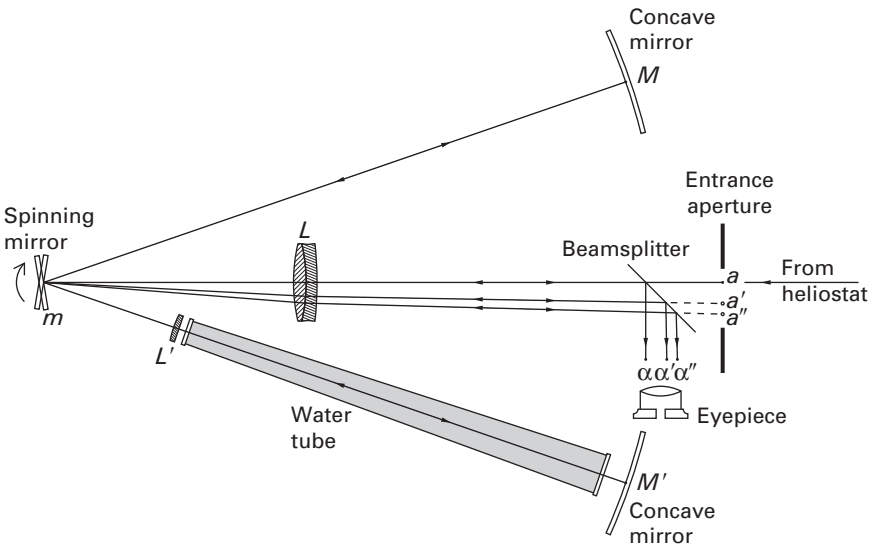


Figure 3.1 Foucault's experimental arrangement. (From Tobin, 2003, 125. Image courtesy of William Tobin.)

spherical mirrors M and M'. An additional lens L' and a 3 m tube of water are placed along the beam path between m and M' (the lens L' corrects for refraction due to the water in the tube). The mirrors M and M' then reflect the image of the wire in the entrance aperture back to m, through L and to a glass plate (a 'beamsplitter') that deflected the image to an eyepiece. With m stationary and oriented either towards M or M', a rule was marked in the eyepiece to indicate the original position of the image of the wire.

When m is set spinning, however, given the finite velocity of light, it rotates through some small angle during the time it takes light to make the journey from m to M or M'. As a consequence, the image of the wire is shifted to the right in the eyepiece.[3] Foucault was able to compare the shifts of the image from M and M' simultaneously in the eyepiece. The image from M' was shifted further to the right, indicating that m had rotated further during the time it took the light to make the round trip from m to M'. Foucault concluded that light had traveled more slowly along the water-path than along the air-path.

[3] Quite apart from the comparison between the air-path and the water-path, this enables one to use the apparatus to measure the speed of light. But the short distance between m and M needed to maintain a bright and clear image in Foucault's set up limited the accuracy of this measurement. Michelson carried out a variation on Foucault's experiment with a different placement of the lens, a much greater separation of the mirrors, and hence greater accuracy (Michelson, 1879).

The verdict that is often rendered from these results is that they showed the falsehood of Newton's corpuscular theory of light. One author describes Foucault's response to the experiment thus: "The emission theory was dead, incontestably incompatible with the experimental results!" (Tobin, 2003, 127). Here, however, Duhem cautions us to make an important distinction: The experiment does not show that a particular claim such as "light consists of corpuscles" is false. "[W]hat the experiment declares stained with error is . . . the whole theory from which we deduce the relation between the index of refraction and the velocity of light in various media" (Duhem, 1991, 187). That is, the prediction that light should travel faster in water than in air depends not only on the premise that light consists of corpuscles emitted by luminous bodies, but also on additional assumptions, such as those about the forces exerted on such corpuscles by matter of a refracting medium and the reactions of the corpuscles to such forces. The result that light travels more slowly in water than in air logically demands only that we acknowledge that not all of those premises are true. This, however, does not allow us to judge the truth or falsehood of the premises individually. "In sum," Duhem writes, "the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses," and if the group fails the test, "the experiment does not designate which one should be changed" (Duhem, 1991, 187).

3.2.2 Duhem on physics

Duhem begins his discussion with a contrast between physiology and physics, noting with approval the advice of physiologist Claude Bernard that for the physiological investigator "theory should remain waiting, under strict orders to stay outside the door of the laboratory" (Duhem, 1991, 181). In a physics experiment, however, Duhem notes that "it is impossible to leave outside the laboratory door the theory that we wish to test, for without theory it is impossible to regulate a single instrument or to interpret a single reading" (Duhem, 1991, 182). Duhem concludes this from the fact that the theories of physics themselves describe the operation of instruments used by physicists to make measurements, produce phenomena, and manipulate entities and processes.

Because the use of instruments such as thermometers, oscilloscopes, phototubes, etc., and the interpretation of the data resulting from these devices

depend upon the theories of physics that explain how such devices work, performing an experiment in physics requires one to assume from the outset a substantial body of physical theory. He acknowledges that experimenters working in chemistry and physiology use some of these same instruments, and thus rely on similar theoretical assumptions, but in these cases the theory assumed is concerned with a domain distinct from that of the theory being tested. Only the physicist is "obliged to trust his own theoretical ideas or those of his fellow-physicists" (Duhem, 1991, 183).

Duhem here makes an interesting concession: "From the standpoint of logic, the difference is of little importance; for the physiologist and chemist as well as for the physicist, the statement of the result of an experiment implies, in general, an act of faith in a whole group of theories" (Duhem, 1991, 183). As we continue our discussion, we will return to the question of whether "an act of faith" is an apt description of what is needed from the experimenter. Setting that issue aside, it might seem that Duhem here admits that his distinction between the implications of underdetermination for physics as opposed to other fields is of little importance. As we will see, however, the qualifying expression "from the standpoint of logic" is significant. Duhem does not think that the standpoint of logic alone matters for judging the import of an experiment.

3.2.3 Duhem's problem and "good sense"

As we noted above, Duhem's thesis UD is unobjectionable. For the scientific investigator, however, UD cannot be the end of the discussion because the thesis burdens the investigator with a pressing *problem* that is both epistemological and practical. At some point one has to report what one has learned from the experiment performed. What should you say? If you report only that some one of a whole set of premises used to derive a predicted result has been shown (or not been shown) to be false, the very next question you will be asked is, "Can you say anything about *which* premises we ought to at least suspect of being incorrect?"[4] You would like to answer positively, but logic

[4] Should the outcome of the experiment be consistent with the prediction from the hypothesis under test, one might be asked a somewhat different question: "Were that hypothesis false, is it likely one would have gotten a result that did not agree so well with the predicted result?"

does not give you any basis for doing so. Call this the 'underdetermination problem,' or, if you like, 'Duhem's problem.'

Duhem notes that different responses to any given underdetermination problem may be not only possible but even defensible in some sense. One physicist, having encountered an experimental result at odds with the prediction from some theory, might choose to "safeguard certain fundamental hypotheses" in the theory by making adjustments to some of the less central assumptions (about causes of error, or about how the instruments work). Another, contemplating the very same result, may take it as counting against those same fundamental hypotheses protected by the first physicist (Duhem, 1991, 216–217). Duhem notes of these two physicists that "The methods they follow are justifiable only by experiment, and if they both succeed in satisfying the requirements of experiment each is logically permitted to declare himself content with the work that he has accomplished" (Duhem, 1991, 217).

Let us note two aspects of this statement. First, Duhem does require that each physicist must satisfy "the requirements of experiment." Although he does not elaborate on this point, this presumably means that one may not wantonly disregard the fact that certain kinds of error need to be controlled for, that measurements need to be made following a consistent procedure, that sources of bias in one's sampling procedure need to be taken account of, etc. Second, both physicists are "logically permitted" to declare themselves satisfied. But, as Duhem notes immediately following this passage, "Pure logic is not the only rule for our judgments; certain opinions, which do not fall under the hammer of the principle of contradiction are in any case perfectly unreasonable." Duhem attributes such judgments of unreasonableness to a faculty that he calls "good sense" (Duhem, 1991, 217).

By his own admission, Duhem's invocation of good sense as the means by which to solve underdetermination problems lacks precision and clarity. The reasons for preferring one judgment to another that good sense provides are "vague and uncertain"; they are apt to be a matter of (temporary) disagreement among different investigators, leading to controversies. But for precisely this latter reason, scientists have an obligation to *cultivate* their good sense, for by doing so they may shorten the time it takes for an inadequate hypothesis to be replaced by "a more fruitful" one (Duhem, 1991, 218). Here it is evident that for Duhem the underdetermination problem should not cause us to despair of the possibility of making progress in science, for good sense (whatever that might really be) will *eventually* produce, not simply the

termination of disagreement, but agreement on what is in some sense the *right* decision. "[I]n order to estimate *correctly* the agreement of a physical theory with the facts, it is not enough to be a good mathematician and skillful experimenter; one must also be an impartial and faithful judge" (Duhem, 1991, 218; emphasis added).

3.3 Underdetermination revisited: Quine

Duhem's discussion fails to provide any very concrete advice regarding "good sense." Aside from being impartial, can experimenters *do* anything that would help sharpen their experiments to point more definitively to a particular conclusion?

Although one might think that making progress on this kind of question would be the natural next step, discussion of Duhem's thesis took a different direction in the mid twentieth century. In his influential 1951 essay "Two Dogmas of Empiricism," W. V. O. Quine took up the UD thesis and put it to a use quite different from Duhem's (Quine, 1980). But, because he cited Duhem in a footnote as having defended the doctrine that "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body" (Quine, 1980, 41), readers came to associate Duhem's and Quine's views. The thesis Quine defended came to be known as "the Duhem–Quine thesis."[5]

Unlike Duhem, Quine was not primarily concerned with the role of logic and method in scientific experiment. Instead, Quine sought to address issues concerning language. In particular, Quine's argument attacked two widely held philosophical views about the meanings of words. The first of the two "dogmas" that he rejected holds that there is a fundamental and important distinction between *analytic* statements (such as "every carnivore eats meat") that are true or false in virtue only of the meanings of the words used to express them, and *synthetic* statements (such as "an emu is wearing my favorite scarf") the truth or falsehood of which depends on

[5] This conflation is perhaps not so surprising. It seems likely that Quine's audience, like Quine himself, had not read Duhem (the reference is absent from the original version (Quine, 1951)), being more acquainted with the Anglo-American variant of logical empiricism to which Quine was responding. See (Quine, 1991, 269) for his own account of how he came to insert the reference to Duhem and (Gillies, 1993) for a useful comparison of Duhem and Quine.

something more than the meanings of the words used to express them. The second dogma, which he calls *reductionism*, Quine describes as "the belief that each meaningful statement is equivalent to some logical construct upon terms which refer to immediate experience" (Quine, 1980, 20). (In the next chapter we will consider a view akin to what Quine calls reductionism in our discussion of logical empiricism.)

Let us bypass the arguments upon which Quine bases his rejection of these two dogmas, and turn our attention to the conclusion that he draws. He claims that abandoning reductionism and the analytic/synthetic distinction leads one to the view that our knowledge, taken all together, "is like a field of force whose boundary conditions are experience" (Quine, 1980, 42). Experience, that is, constrains our beliefs about the world in some way, but only weakly. If the expectations we form about future experiences, based on the beliefs that make up the "field" are in conflict with our actual experience, we have to change our beliefs somehow. "But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience" (Quine, 1980, 42).

How much latitude, you ask? Quine's answer is striking: "Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system." Moreover, we can make any adjustment we like because "no statement is immune to revision," not even the laws of logic (Quine, 1980, 43). As Quine well knows, we do not make or withhold revisions arbitrarily. We have *reasons* (in the appropriate circumstances) for taking our experience of strawberry flavor to result from our having eaten an actual strawberry rather than from our having a strawberry-hallucination or being possessed by strawberry demons. But Quine insists that these reasons are ultimately "pragmatic." It would be difficult and costly to revise our beliefs so that strawberry experiences were to be interpreted by a new set of beliefs about strawberry-experience-inducing-demons. Nonetheless, we *could* do so, and the fact that we do not reflects simply our reluctance to undertake such costly revisions to our beliefs when they are so easily avoided.

3.4 A big deal?

A comparison of Duhem's and Quine's discussions reveals that, although they share the thesis UD as their premise, they conclude rather different lessons

for the scientific enterprise. According to Quine, were it not for "pragmatic" considerations, our reasoning about the world would be unconstrained, since even "statements of the kind called logical laws" may be revised. Duhem draws a somewhat more restrained conclusion, noting simply that quite different responses to an experimental outcome can be properly defended, provided that the investigators have met the "requirements of experiment."

Should we take the implications of UD to be a big deal? That human reasoning involves more than logic alone is hardly a surprise. As a purely formal science concerned with deductive relationships between one statement and another, logic cannot by itself solve the problem of differentiating between that which is deserving of belief and that which is not. But, as Duhem notes in a previously quoted passage, "certain opinions, which do not fall under the hammer of the principle of contradiction are in any case perfectly unreasonable" (Duhem, 1991, 217). The pressing question is not whether it is logically possible to hold any statement true "come what may," but what standards beyond logical consistency we should apply when deciding what to believe and what status those standards have, given that they are not derivable from logic alone.

Larry Laudan has argued this point, noting that Quine's claim that "any statement can be held true come what may" sounds a lot less methodologically significant if we recall that the word "can" expresses different kinds of possibility depending on how it is used (Laudan, 1996). In the present context we could understand Quine's claim to mean "given any particular statement that contradicts other beliefs a person holds there will be a strategy for maintaining belief in that statement without such contradiction by modifying other beliefs." This does not, however, mean that one may *justifiably* or *rationally* believe a statement regardless of the experience that one has had.[6] Deciding what one *ought* to believe demands more than the answer to the question of what one *can* believe while remaining logically consistent. (Alternatively, we might say, the question for philosophy of science might not really be about what one ought to *believe* at all, but rather what theory is best-corroborated, -supported, or -tested. Popper's falsificationism is

[6] As Laudan goes on to argue, however, one finds precisely these latter claims defended on the grounds of the **UD** thesis, including the claim that "Any theory can be shown to be as well supported by any evidence as any of its known rivals" (Laudan, 1996, 39).

one philosophy of science in which questions about belief are not central to scientific method. We will encounter another such approach in the error-statistical philosophy of Chapter 9.)

Consider Foucault's observation that the wire-image produced by light traveling the air-path was shifted less than that resulting from the water-path. Foucault confronted an underdetermination problem: whether to present his results as counting against the view of light as consisting of corpuscles or as conflicting with some other assumption, since the result refutes not any isolated hypothesis about the nature of light but rather "the whole theory from which we deduce the relation between the index of refraction and the velocity of light in various media" (Duhem, 1991, 187). In truth, as discussions of underdetermination in the wake of Quine's essay have noted, the situation is even worse, since it is not only theoretical assumptions that are needed to logically connect the corpuscular theory to the predicted shift; also needed are statements describing the set up of the experiment. For example, were the distance from mirror m to mirror M sufficiently *greater* than the distance from m to M', then the result obtained by Foucault might come about even were the corpuscular theory to be correct.

Consequently, had Foucault been motivated to protect the emission theory, he could have done so in a logically consistent way, simply by denying that the two beams of light traveled the same distance in the manner required for the experiment to be probative. Even had Foucault personally measured the two paths to be the same length, he could have insisted that the measuring device had changed length during the measuring process, or that he had been under the influence of powerful hallucinogenic drugs while recording the result of that measurement. (Remember from the last chapter Popper's point that logic does not apply directly to experiences, but only to statements that describe experiences.) These possibilities were also open to members of the scientific community who heard reports of Foucault's experiment, including those who, unlike Foucault, actually *were* motivated to protect Newton's theory from refutation.

He *could* have done any of these things, but it seems reasonable to expect that Foucault would have encountered serious criticism for such a procedure. Failing to ensure that the light-path and air-path are of equal length would constitute a serious defect in experimental design. Certainly this would be covered by Duhem's phrase "the requirements of experiment"! Other changes in belief (pleading hallucination, assuming some change in the measuring

device), apparently motived *only* by a desire to protect the emission theory from falsification, seem, in Duhem's words "perfectly unreasonable." Moreover, describing Foucault's decision *not* to respond to his results in these ways as an "act of faith" (as Duhem does) seems misleading if by that we mean that he could not have had *reasons* (beyond convenience) for relying on the assumptions that he did. Acquiring such reasons would seem to result from, or perhaps even be the point of, careful safeguards aimed at securing the reliability of one's experimental procedure.

3.4.1 An epistemology of experiment

The strategies that experimentalists use to validate their conclusions are at the core of an *epistemology of experiment* advocated by Allan Franklin, based on his detailed studies of a wide variety of experiments in physics (Franklin, 1986, ch. 6; Franklin, 2012).[7] Although a detailed discussion of Franklin's strategies would take us away from the focus of the present chapter, they include the use of calibration techniques (in which known phenomena are reproduced to test the functioning of an apparatus), reliance on independently well-tested theory (either to explain the phenomenon produced in the experiment or to warrant reliance on an apparatus whose proper operation that theory predicts), the use of statistical arguments, and, perhaps most importantly for our present concern, the elimination of plausible sources of error and alternative explanations.

Although the logical core of the Duhem problem stands, these strategies provide important resources for scientists to resolve underdetermination problems in a piecemeal, yet reasoned, way. Even the conscientious application of them does not guarantee that one will not fall into error in one's conclusions, since they draw on resources beyond the truth-preserving strategy of deductive inference. In particular, they require investigators to rely on *judgments* of various kinds, such as the extent to which alternative explanations are plausible, or whether a theory has been sufficiently well-tested to make its application in a particular instance trustworthy.

[7] Although Franklin's strategies were drawn from studies of experiments in physics, David Rudge has applied them to an experiment in biology investigating industrial melanism (Rudge, 1998, 2001), and Franklin himself applied them to a discussion of Meselson and Stahl's experimental work on DNA replication (Franklin, 2005, ch. 3).

3.5 Conclusion

We have seen how both inductivist and falsificationist ways of thinking about scientific reasoning face the problem of a logical gap. What is less clear, however, is the significance of that gap for the results of scientific inquiry. If logic exhausted our resources for evaluating scientific theories, it would seem to follow that our decision to respond in one way rather than another to our observations of the world would ultimately be arbitrary, or at best driven by mere preference based on considerations of factors such as convenience. Something like this latter view does seem to emerge from Quine's discussion of underdetermination in "Two dogmas of empiricism."

But Duhem's discussion should give us pause before conceding the apparent assumption that we have only logic (and considerations of convenience) to go on when we have to decide what to make of our observations. Duhem points, if perhaps vaguely, to two additional resources: (1) methodology ("the requirements of experiment"), which provides standards for the competent execution and recording of experiments or observations, so as to reduce our susceptibility to error and enhance our ability to learn about the hypotheses that we subject to experimental tests, and (2) judgment ("good sense"), grounded in prior study of both phenomena and theories within a domain, which, we might suppose, allows the investigator to judge the extent to which a particular theory can be regarded as a plausible contender in light of all the relevant knowledge. To be sure, we would like to have a better, more clearly articulated explanation of these resources, and Franklin's epistemology of experiment represents one effort in that direction. We will consider further efforts involving *probability* in Chapters 8 and 9.

4 Logical empiricism and scientific theories

4.1 Introduction

We have seen how Duhem's arguments indicate limitations on how far deductive logic alone can guide us to decisions about scientific theories based on experimental outcomes. In the early decades of the twentieth century, another group of philosophers and scientists on the European continent were pursuing a philosophical program that attempted to exploit the resources of formal logic as far as possible in the articulation of a 'scientific philosophy.'

In this chapter we will take a brief look at the philosophical movement known as logical empiricism or logical positivism, paying particular attention to a central project of the movement: an account of the structure of scientific theories. The preoccupation of philosophers of science with the question of theory structure has its roots in an underlying commitment – rarely examined or even articulated throughout most of the twentieth century – to the importance of theories as the primary contribution of scientific inquiry to human knowledge. Later, some philosophers challenged this assumption, arguing that experimentation produces knowledge that is philosophically significant beyond simply its role in testing broad explanatory theories (Ackermann, 1985; Franklin, 1986; Hacking, 1983; Mayo, 1996). Nonetheless, the question of theory structure remains highly relevant to our understanding of scientific knowledge, and the fate of logical empiricist views has had important consequences for the subsequent development of philosophy of science, as we will explore in the next several chapters.

4.2 The Vienna Circle and logical empiricism

4.2.1 The Vienna Circle and philosophy of science

Although the questions pursued by contemporary philosophers of science have roots deep in the beginnings of Western philosophy, the emergence of philosophy of science as a distinct discipline within philosophy, with its own professional organizations, its own scholarly journals, and its own programs of specialized training, was a twentieth-century phenomenon. Although a thorough discussion of that emergence in its institutional, cultural, and intellectual dimensions goes beyond our present concerns, it is safe to say that integral to the formation of philosophy of science as a specialized field of study were the activities and eventual dispersion of a group of philosophers and scientists who met in Vienna, Austria and have been loosely identified as 'the Vienna Circle.' In addition to fostering the development of a distinctive philosophical approach through their interactions with each other, the Vienna Circle influenced philosophers from other countries who came to study and discuss issues with them. Their influence became even greater when the Nazis seized power in Germany and Austria, at which time many members of the group (which included quite a few Jews and leftist intellectuals who correctly perceived the Nazis as an imminent threat) fled to other countries, many of them making their way, eventually, to the United States, where they became influential scholars and teachers.

Recent scholarship has shown that the philosophical views of those affiliated with the Vienna Circle were quite diverse (see, e.g, Friedman, 1999; Richardson & Uebel, 2007; Stadler, 2003). Nonetheless one can identify as central to Vienna Circle philosophy a broad idea about how philosophy might articulate a scientific way of knowing that would be free of bias, free of dogma and superstition, free of nationalist ideology, and beneficial to humankind. Inspired (like their fellow Viennese Karl Popper, with whom they vehemently disagreed on many issues) by recent advances in physics, they sought to achieve these lofty goals with the use of recent advances in formal logic and with the guidance of an idea about what separated meaningful from meaningless discourse.

4.2.2 Meaning vs. metaphysics

The idea that played such an important role in the Vienna Circle held that meaningful discourse was distinguished from meaningless nonsense by the

fact that the former, but not the latter consisted of statements that could be *verified by experience*. This idea came to be known as the verification principle and the philosophical orientation in which it played a central role came to be called either *logical positivism* or *logical empiricism*. The verification principle was not merely stipulated, but arose from the belief that the clearest exemplars of meaningful discourse are to be found in such successful empirical sciences as physics, while the history of metaphysical theorizing from the ancient Greeks onwards provides plenty of examples of meaningless statements that could never be verified by any observation of any kind.

While physical science provided the Vienna Circle with their best example of a knowledge-generating enterprise, it also proved to be the source of their greatest challenge. It is all fine and well to *say* that the scientific discourse of physics can be seen to be meaningful by the fact it consists of statements that can be verified by experience, but is that really true? For example, Einstein, writing in 1918, explained that he considered his General Theory of Relativity (regarded by logical empiricists as a signature achievement of modern science) to be based on principles entailing that the "G-field" (the "condition of space" described by a mathematical object called the "fundamental tensor," essential to the articulation and application of Einstein's theory) "is determined *without residue* by the masses of bodies" (Einstein, 1918, 241–242, quoted in J. D. Norton, 1993, 306; emphasis in original).

To those not familiar with General Relativity, this statement sounds opaque at best. Even for those well-versed in the relevant physics, however, this statement does not obviously show itself to be a statement that can be verified by experience as required by the logical empiricists' principle, since the G-field itself is not directly observable. To put the problem succinctly, logical empiricists had to show that one could articulate the verification principle in such a way that the examples constituting, by their own reckoning, the clearest cases of meaningful discourse did actually *satisfy* the principle.

The most influential efforts by logical positivists to tackle this problem invoked a linguistic distinction. The terminology employed by a scientific theory would be divided into three distinct categories: (1) the *logical* vocabulary V_L (consisting of terms such as 'every,' 'some,' or 'not'), (2) the *observational* vocabulary V_O, and (3) the *theoretical* vocabulary V_T. Proposals for drawing the distinction between V_O and V_T varied somewhat, but the basic idea was that the truth or falsehood of a statement containing only terms in V_O (terms such as, perhaps, 'red' or 'to the left of' or 'coinciding with') could be decided

directly by appeal to observations, while statements using theoretical terms (such as 'electrically charged,' 'acidic,' or 'G-field') could not be. To vindicate the verification principle, the logical positivists sought to show that in successful empirical sciences, statements that include terms from V_T (like 'this sample is acidic') stand in some *appropriate relation* to statements using only terms from V_O (perhaps 'if this sample is applied to litmus paper, the litmus paper will turn red'). Statements expressing these relations are *correspondence rules* because they indicate what, in experience, corresponds to the theoretical terms invoked.

One problem was how to characterize that relation. It would have been nice – for the logical empiricists – to be able to say that statements in the successful disciplines of empirical science that use theoretical vocabulary can simply be *defined explicitly and completely* with statements using only observational terms. This would require that for every such theoretical statement there is some purely observational statement to which it is logically equivalent.[1] Thus, one might wish to be able to secure the meaningfulness of an assertion using the term 'acidic' by means of a *definition* such as 'this sample is acidic if and only if it is the case that if this sample is applied to litmus paper the litmus paper will turn red.' Such sentences have been called *reduction sentences* because they purport to show how statements involving theoretical terms can be reduced to statements involving only observational terms, which is to say that theoretical terms in the original statement could be replaced by constructions involving only observational terms without losing or changing the meaning of the original.

But, as the logical empiricists well knew, this will not work. First, there is the problem that even a seemingly simple statement such as 'this sample is acidic' appears to mean much *more* than is captured by any one observation-language statement such as 'if this sample is applied to litmus paper, the litmus paper will turn red' (assuming – dubiously – that we can include 'litmus paper' in our observational vocabulary). Being acidic has implications for many different kinds of chemical reactions into which the sample might enter, and not simply its interaction with litmus paper. One might counter this objection by insisting that 'this sample is acidic' needs to be analyzed using some more complicated observational-vocabulary statement describing

[1] Two sentences S_1 and S_2 are *logically equivalent* just in case S_1 logically entails S_2 and S_2 logically entails S_1.

other observational tests of acidity as well as the litmus-paper test, but the defender of reduction sentences clearly now bears the burden of showing that such a reduction can be completed.

That brings us to a second and more fundamental problem. As Rudolf Carnap, arguably the most prominent and influential of the Vienna Circle logical empiricists, acknowledged (Carnap, 1936, 1937), a statement like 'this sample is acidic' *cannot* be equivalent to some statement of the form 'if this sample is subjected to conditions x, then observable condition y will be realized' simply because the latter will inevitably be true in some circumstances in which the former is false. As we noted in Chapter 1, a conditional statement of the form 'if p then q' is considered by logicians to be true whenever p is false, regardless of whether or not q is true (in the case where p is true, 'if p then q' is true in case q is also true, false otherwise). For the present example, this has the consequence that 'if this sample is applied to litmus paper, the litmus paper will turn red' is true for any sample that is not applied to litmus paper, regardless of its chemical properties or composition. Thus it will turn out that many things which are certainly not acidic (the carton of milk in my refrigerator, for one) will make the conditional 'if the sample is applied to litmus paper, the litmus paper will turn red' come out true, simply in virtue of never being applied to litmus paper.

Logical positivists sought to get around this problem by weakening the relationship between statements using V_T and those using V_O, but struggled to avoid weakening the requirement so far as to allow 'metaphysical' statements (which they had intended to eliminate as meaningless) to satisfy the criteria of cognitive meaningfulness thus developed.

Here is an example of such an effort: Given that asserting the logical equivalence of a statement including vocabulary from V_T with a statement using terms only from V_O and V_L proved problematic, one might use, instead of a definition sentence asserting logical equivalence, a sentence of a different form. Compare the following two sentences:

1 'This sample is acidic if and only if, if it is applied to litmus paper, then the litmus paper turns red.'
2 'If this sample is applied to litmus paper, then it is acidic if and only if the litmus paper turns red.'

Sentence (1) is a reduction sentence asserting an explicit definition of 'acidic' that falls prey to the objection under consideration. Sentence (2), however,

makes no positive or negative commitment regarding the acidity of samples that are never applied to litmus paper, and thus falls short of an explicit definition of acidity in general; yet, it does establish a logical connection of a sort between the theoretical term 'acidic' and the observational (or at any rate less abstract) terms 'litmus paper,' 'applied to,' and 'red.' Such more 'liberal' developments in the articulation of the logical empiricist approach to scientific theorizing were prominent through the middle decades of the twentieth century (Carnap, 1936, 1937, 1955; Hempel and Oppenheim, 1948; see the introduction to Suppe, 1977 for an overview).

One more important modification is worth noting. Although statements about particular samples and slips of litmus paper might be verifiable, scientific discourse, as we have discussed in previous chapters, depends in large part upon statements of generalizations, and these clearly cannot be verified, if by that we mean definitively proven to be true. (Carnap, in fact, went further and affirmed that no synthetic statement at all could be "verified completely and definitively" (Carnap, 1936, 427).) The verification principle thus needed to be, and was, re-articulated in terms of some weaker notion, typically *confirmation*. Such efforts led naturally to a considerable body of technical work on confirmation theory, which philosophers of science have continued to pursue even in the absence of any ongoing explicit commitment to the original project of the logical empiricist movement.

4.3 From sentences to models

In the next two chapters we will consider some of the developments of the 1960s and 1970s that were seen as marking the demise of the logical empiricist movement. Before we turn to these, though, it is worth considering a challenge to a central assumption of the logical empiricists regarding what kind of thing a scientific theory is.

One point on which all of the logical empiricists (and their falsificationist foes) agreed, was that a scientific theory consists of a structured set of *sentences*. It is precisely this feature that enables us to bring the apparatus of formal logic to bear *directly* on the theory.

The logical empiricists thus hold a *syntactic* conception of scientific theories ("syntactic" because of its emphasis on the construction and ordering of sentences out of their elements, i.e. the *syntax* of sentences). More precisely, they consider a theory to consist of a set of sentences with a certain logical

structure. First, one should identify some (hopefully finite) number of sentences that use a theoretical vocabulary to articulate the basic principles of the theory. These are the *axioms*. Then one identifies the *correspondence rules* that relate the theoretical terms in the axioms to terms in the observational language. By means of the correspondence rules one can then deduce from the axioms the observational implications of the theory. The theory then consists of the axioms, the correspondence rules, and all of the other sentences that can be deduced from them and their deductive consequences.

Part of the decline of the logical empiricist approach has involved a partial turn away from the syntactic view as a number of philosophers of science have embraced some version of a *semantic view* of scientific theories (Sneed, 1971; Stegmüller, 1976; Suppes, 2002; van Fraassen, 1972). Whereas syntax is concerned with the structure of sentences, semantics is concerned with relations between sentences, or language more broadly, and whatever it is that we use language to talk about; semantics thus is concerned with such relations as meaning, reference, and representation.

The semantic approach to theories holds that a theory is not a collection of sentences at all, but rather a collection of *models*. So the sentences scientists use when engaging in theoretical discourse do not *constitute* the theory, but instead *describe* the theory in the sense that a sentence describes the structures that are its models. If we wish, therefore, to understand what a scientific theory is, we should therefore look to the models, rather than the sentences. As stated by Suppe, "Theories are extralinguistic entities which can be described by their linguistic formulations. The propositions in a formulation of a theory thus provide true descriptions of the theory, and so the theory qualifies as a model for each of its formulations" (Suppe, 1977, 222).

The term 'model' has a very diverse range of meanings in science, which can make a unique and precise characterization of the semantic view somewhat elusive. In the appendix to this chapter, I introduce a rather technical understanding of models, according to which a model is *a structure in which sentences with a particular logical form are true*. Scientists, however, use a wide variety of means for describing the models that lie at the heart of their theorizing. Among these means, mathematics is prominent, but scientists also use visual representations (schematic drawings, flow charts, etc.) and three-dimensional models (such as the stick-and-ball models used for depicting molecular structures); increasingly, computer-based simulations are used as models (Parker, 2009; Winsberg, 2010). The term 'model' is indeed a very

broad term, and not all of its uses fit neatly the 'models-as-structures' idea just articulated (see Achinstein, 1968 for a discussion of different kinds of models in science). In a broader sense, we can think of the turn (by many philosophers of science, anyway) against the syntactic conception of theories as embracing the idea that we can better understand scientific theories by looking at the structures of theoretical models (in whatever form they take) and the ways that those models are connected (often with the help of additional 'mediating models') with the data obtained by experiment and observation (Morgan & Morrison, 1999).

Still, where the notion of a model as a structure in which sentences with a particular logical form are true *does* apply, we enjoy the benefit that we can apply the resources of formal logic, not directly to the theory, as the logical empiricists wished, but *indirectly*, to the sentences that describe those theoretical models.

4.3.1 Models as idealized structures

We must note an important nuance in how such structures relate to the world as we experience it. Advocates of the semantic view emphasize the fact that the application of a theory always involves the introduction of some kind of *idealization* regarding the phenomena or systems to which we wish to apply the theory, and that it is these idealized structures that are described by our linguistic formulations in science.

To see how this works, consider an example from elementary chemistry. Chemists sometimes invoke Charles' Law, which states that at constant pressure, the volume of a given mass of an ideal gas is directly proportional to its absolute temperature. Sometimes the law is carelessly formulated as "the volume of a confined gas is directly proportional to its temperature in degrees Kelvin." Note one glaring difference between these two formulations. The first, more correct statement, refers to an "ideal gas at constant pressure" while the latter simply refers to "a confined gas." The distinction is significant! An ideal gas is really a fiction. It is composed of molecules that take up no space at all, that do not lose any kinetic energy when they bounce into one another, and that have no effects on each other's motion except when they collide (they exert no attractive or repulsive forces upon one another). None of these facts about ideal gases is true of any actual gas, but in many cases

we can get a reasonably good approximation of the behavior of a real gas by applying Charles' Law *as if* these assumptions were true.

Although the idealization is hard to miss in the case of Charles' Law (the word "ideal" occurs in the statement of the law itself!), the same point holds true in various ways and to various degrees wherever we see scientists applying theories. If we wish, for example, to apply Newton's laws of motion to predict the motion in free fall of an object near the surface of the Earth, we might begin by assuming that the object has all of its mass concentrated at a single point. This not only simplifies the task of describing how the Earth's gravitational field acts upon that object, it also removes the complication of the effects of air resistance. The consequences of such a simplification will depend on the problem at hand. If the object is a bowling ball, our idealization will be less damaging to our predictions than if it is a feather.

In any such case we can partially 'de-idealize' our description by making somewhat more realistic assumptions; we might try to give a more realistic description of the distribution of mass in the body that has been dropped. The price we will pay is that our attempt to apply the theory will become more difficult. The important point about all of this for our present purposes is that, for advocates of the semantic approach to theories, the models that constitute a given scientific theory are not the world as we experience it or even some part of that world, but instead are *idealized structures that are described by the linguistic formulations associated with that theory*.

It is worth pausing for a moment to contemplate this point. Before we begin to think philosophically about science, I think most of us are inclined to suppose that what scientific theories do is simply to *describe what, in general, the world is like*. For example, if asked "What does Charles' Law describe?" we would be inclined to say something like "It describes the relation between volume and temperature for gases at fixed pressure." The position under consideration here insists that the linguistic formulations we use to describe Charles' Law describe idealized structures that are not really part of the world at all.

Here you might think that if the semantic view leads us to have to abandon the idea that scientific theories describe the real world rather than our fictional, idealized simulacra, then so much the worse for the semantic view. Perhaps we should stick to the idea that theories should be understood in terms of sentences that either directly describe the world or from which we can logically deduce sentences that do. However, sticking to the syntactic

view of theories will not solve this particular problem. The problem of idealization remains, but we would now formulate it differently: The sentence expressing Charles' Law is not true as a description of any *actual* gas, but only of non-existent, ideal gases; the sentences we use to make predictions using Newton's mechanics do not describe the motions of any actual bodies, but only of idealized 'point masses'; and so on. At what point do our scientific theories make contact with the real world?

Whether we think of theories in semantic or syntactic terms, there is a question about how, and even whether, scientific theories could be regarded as describing or representing the world as it really is. Defenders of both approaches have suggested ways of tackling this problem. For example, some defenders of the semantic approach have argued that, although the linguistic formulations associated with a theory describe idealized structures, those structures themselves can be regarded as *resembling*, in certain respects, those aspects of the world that the theory is developed in order to understand or explain (Giere, 1999).[2] The most common response to the problem of idealization on behalf of a realist understanding of syntactically conceived theories is to invoke the concept of *approximate truth*: Although even our best theories may not be precisely true of the world as stated, they are approximately true. Both of these responses have to confront substantial objections, and some philosophers of science have responded by concluding that one should not, after all, think of scientific theories as telling us what the world is really like at the most basic and possibly unobservable level.

Further consideration of these issues will have to await Chapter 10, in which we discuss scientific realism and anti-realism.

4.4 Conclusion

The logical empiricists, while far from being the first philosophers to engage problems related to scientific knowledge and theorizing, were historically significant in their efforts to articulate a far-reaching approach to science. They drew upon the formal apparatus of logic to serve a broader philosophical

[2] This is still a bit too simple. The resemblance relation really should be thought of as holding between the *theoretical* model and *another* model, based on data (i.e., measurements of real-world things or processes) that is sometimes called a *model of the data* or an *empirical model* (Suppes, 1962).

agenda. Much of the discourse as well as professional infrastructure of current philosophy of science owes some kind of debt to the Vienna Circle and those it engaged as logical empiricism came to dominate philosophy of science in the English-speaking world in the middle of the twentieth century. The semantic approach to theories presented a significant challenge to logical empiricism, in that it challenged the basic framework in which logical empiricists pursued their project of articulating just how scientific theories can remain empirically meaningful while introducing theoretical terms that do not directly refer to anything observable.

Nonetheless, the challenge to logical empiricism from the semantic approach is a *narrowly focused* challenge insofar as it does not directly confront the broader aims of logical empiricism, viz., to articulate a 'scientific philosophy' in which disagreements over matters of substance could, to at least a significant extent, be treated as susceptible to resolution by appeal to the neutral ground of empirical observation. Moreover, the semantic approach, like logical empiricism, originated with attempts to employ formal techniques to elucidate philosophical problems.

In the next three chapters, we will consider a more sweeping challenge to the approach taken by logical empiricists to science. We will see how, drawing widely on the history of science (as well as philosophical analysis), the philosophers Thomas Kuhn, Imre Lakatos, and Paul Feyerabend, in their own distinctive ways, reoriented the attention of philosophers and challenged some of the central tenets of logical empiricism, with important consequences for the subsequent development of philosophy of science.

4.5 Appendix: models as structures

A thorough understanding of the semantic view of theories requires some discussion of semantics.[3] In particular we are interested in a general characterization of the relationship between sentences having a certain logical form and the things that make such sentences true. Our first order of business

[3] What follows is a rather rough-hewn outline of some important concepts that are well worth spending time studying in greater depth and with greater care. There are many good logic texts that introduce these notions. I use (Barker-Plummer, Barwise, & Etchemendy, 2011) when I teach this subject. A more advanced but still introductory text is (Enderton, 2001).

in this subsection will be to look a little more carefully at the logical form of sentences, distinguishing different elements that we will find there. We can then move on to a characterization of *structures*, and finally arrive at a definition of what it is for a structure to be a *model* of a theory.

Sentences, formulas, satisfaction

Remember Zig and the doughnuts? In Chapter 2 we discussed how the Law of Doughnut Delectability ("All doughnuts are delicious") could be represented in logical form as:

LDD′: For any thing whatsoever, if it is a doughnut, then it is delicious.

This sentence contains (we will suppose, for the sake of argument) the theoretical term 'doughnut' and the observational term 'delicious'. (Perhaps it is a correspondence rule? Never mind. The Vienna Circle would not consider this to be part of science, however much they may have enjoyed doughnuts.) Both of these are what logicians call *predicates*, i.e., they are terms that are in some way descriptive of a thing. They are both *unary* predicates, meaning that they apply to just a single thing. We can create a simple sentence by using a *copula* (like 'is', or some other form of 'to be') to combine a predicate term with the name of an individual (or some other linguistic device for singling out an individual, like the indexical term 'this'): 'R773b is a doughnut'.[4]

We can replace both predicate terms in *LDD′* with *predicate symbols* such as *P* and *Q*, and we will be left with just the logical structure of the sentence, which we call a *sentence schema*:

L_1: For any thing whatsoever, if it is *P*, then it is *Q*.

Now we have a sentence that is no longer about doughnuts. In fact, it is no longer about anything. This same logical expression is what we would arrive at if we had started with the sentence 'All cats are carnivores' or 'All zombies are yoga instructors'. Here is another, slightly more complex sentence schema in our abstract logical language:

L_2: For any thing whatsoever (call it *x*), if *x* is *P*, then there is something (call it *y*), such that *y* is *Q* and *x* stands in relation *R* to *y*.

[4] Perhaps you can think of a better name for a doughnut?

This slightly more complex logical structure is shared by the sentences 'Every dog is owned by a zombie', 'Every electron attracts some positively charged particle', and 'Every doughnut tastes better than some oven mitt'. Because relations like '*x* is owned by *y*', '*x* attracts *y*', and '*x* tastes better than *y*' obtain between pairs of things they are called *binary* predicates. '*x* is between *y* and *z*' is a tertiary (three-place) relation and we can contemplate and express relations obtaining between even larger numbers of individuals. (You might also at this point notice that our use of the pronoun 'it' in L_1, for example, corresponds exactly to the use of a variable like *x*.)

Notice that both L_1 and L_2 begin with 'for any thing whatsoever,' which is then followed by some expression that, as we noted in Chapter 2, is not quite a sentence, like 'if it is *P* then it is *Q*'. We can give this latter expression an interpretation such as 'if it is a doughnut then it is delicious', but this *formula* is still not a sentence because 'it' does not here refer to any individual. Consequently the formula has an indeterminate truth value. It is neither true nor false. The expression 'for any thing whatsoever' is also not a sentence, but a *quantifier*, specifically, the *universal quantifier*. Intuitively, it serves to tell us to what range of things the rest of the expression applies: to each and every thing. Sentence schema L_2 contains another quantifier as well: 'there is something such that'. This *existential quantifier*, again intuitively, indicates that the expression that follows applies to at least one thing (but it does not name the thing or things in question).

Although a formula like 'it is a doughnut' has no truth value, we can turn it into something that does by replacing occurrences of 'it' with the name of some individual: 'R773b is a doughnut'. If the resulting sentence is true (the individual named 'R773b' really is a doughnut and not a bagel or a bassist), then we say that the named individual *satisfies* the formula into which it was substituted.

Structures

A semantics for sentences with quantifiers requires that we consider not just individuals but *structures*. Suppose that we have a language with just two unary predicates *P* and *Q* and one binary predicate *R*. To describe a structure for this language we must specify the following: (1) What constitutes the population of things to which the sentence will be applied? The set of all and only such individuals is called the *domain of discourse D*. (2) Which individuals

Table 4.1 *Structure* $\mathfrak{A}$*: A model of* L_1.

x is *P*	*a* is *P*	*b* is *P*	*c* is not *P*
x is *Q*	*a* is *Q*	*b* is *Q*	*c* is *Q*
x is *R* to *y*	*a* is not *R* to *a*	*a* is not *R* to *b*	*a* is not *R* to *c*
	b is not *R* to *a*	*b* is not *R* to *b*	*b* is not *R* to *c*
	c is not *R* to *a*	*c* is not *R* to *b*	*c* is not *R* to *c*

x in the domain *D* satisfy the formula '*x* is *P*'? Which satisfy the formula '*x* is *Q*'? Which pairs of individuals (x, y) drawn from *D* satisfy the formula '*x* stands in relation *R* to *y*'? The set of individuals or multiples of individuals that satisfy a formula using a single occurrence of a single predicate is called the *extension* of that predicate. Specifying the domain of discourse and extensions of the predicates used in L_1 and L_2 amounts to specifying a *structure* for those sentences.[5]

Note that we can specify a structure without actually identifying the predicates in a sentence with particular words in a natural language, so long as we can attach names to the individuals in the domain. Suppose that we stipulate that there are just four individuals in our domain, named *a*, *b*, *c*, and *d*. We can draw up a table that specifies the extensions of the predicates by listing the *true* simple sentences that can be formed by applying those predicates to particular members or pairs of members of the domain (we will use '*x* is *R* to *y*' to abbreviate '*x* stands in relation *R* to *y*'):

Let's call the structure described in Table 4.1 $\mathfrak{A}$. Here we have a structure that makes the sentence schema L_1 come out true. As careful examination of the table will show, every individual that is *P* is also *Q*. Each individual in $\mathfrak{A}$'s domain of discourse *satisfies* the formula F_1: 'if it is *P* then it is *Q*'. The rule for sentences that begin with the universal quantifier 'for everything *x*' is that they are true for a given structure just in case every individual in that structure's domain of discourse satisfies the formula that follows. However, L_2 comes out false in $\mathfrak{A}$.

[5] Some sentences contain, in addition to predicates like '*x* is *P*' and '*x* stands in relation *R* to *y*', *functions* that refer to one member of the domain of discourse by means of a relation to another member of the domain, such as 'the brother of *x*.' To specify a structure for a sentence that includes functions, one must also specify, for each individual *x* in the domain of that function, the individual that is referred to by applying the function to *x*.

Table 4.2 *Structure $\mathfrak{B}$: A model of L_2.*

x is P	a is P	b is P	c is not P
x is Q	a is not Q	b is Q	c is Q
x is R to y	a is not R to a	a is not R to b	a is not R to c
	b is R to a	b is not R to b	b is not R to c
	c is not R to a	c is R to b	c is not R to c

Here is why: The rule for sentences beginning with 'there is some x such that' is that they are true for a given structure just in case at least one individual in that structure's domain of discourse satisfies the formula that follows. L_2 begins, of course, with 'for everything x'. According to the rule we just learned for that quantifier, L_2 is true in $\mathfrak{A}$ just in case every individual in $\mathfrak{A}$ satisfies the formula F_2: 'if x is P, then there is something y such that y is Q and x stands in relation R to y'. To show that L_2 is false in $\mathfrak{A}$, it thus suffices to find one individual in $\mathfrak{A}$ that fails to satisfy F_2. The individual a is P, according to Table 4.1, so a satisfies the antecedent of the conditional F_2. The consequent of F_2 is 'there is something y such that y is Q and x stands in relation R to y'. Although there are individuals in $\mathfrak{A}$ that are Q, a does not stand in relation R to any of them, so nothing in $\mathfrak{A}$ satisfies the formula 'y is Q and a stands in relation R to y'. Substituting a for x in F_2 therefore yields a sentence that is false in $\mathfrak{A}$. Since a does not satisfy F_2, L_2 is false in $\mathfrak{A}$.

We can specify a different structure (call it $\mathfrak{B}$), using the same domain of discourse, that makes L_2 true but not L_1: In this structure (described in Table 4.2), a is a P that is not Q, meaning that a fails to satisfy 'if it is P then it is Q', so it is not true that every P is Q, but a careful examination of $\mathfrak{B}$ will reveal that there is, for everything that is P, something that is R to that thing that is also Q.

All of this is completely abstract, of course. To see how we can get back to sentences about particular kinds of things and how they are related, try replacing the Ps and Qs and Rs above with suitable English words. For example, let 'x is P' be 'x is a dog' and let 'x is Q' be 'x is a zombie' and let 'x is R to y' be 'x is owned by y'. Now L_2 becomes:

L_2: For any thing whatsoever (call it x), if x is a dog, then there is something (call it y), such that y is a zombie and x is owned by y.

which is another (admittedly quite stilted) way of saying 'Every dog is owned by a zombie'.

Models

We are at last ready to define the term 'model':

Definition 1 *Let* Γ *be a set of sentences, and* $\mathfrak{M}$ *a structure for the sentences in* Γ *(i.e.,* $\mathfrak{M}$ *specifies a domain of discourse, indicates the individuals referred to by all names used in* Γ *and specifies the extensions of all predicates used in* Γ*).* $\mathfrak{M}$ *is a* ***model*** *of* Γ *if and only if every sentence in* Γ *is true for* $\mathfrak{M}$.[6]

It is common, in discussing models in the context of formal semantics, to use the word 'theory' to refer to the sets of sentences that are modeled by certain structures. For obvious reasons, I have not adopted that usage here. In any case, we can regard that as a specialized use of the term 'theory.' What is at issue between syntactic and semantic views of scientific theories is precisely whether this technical use of the word 'theory,' to refer to a set of sentences, adequately reflects what theories are in a way that makes sense of the ways that scientists employ them.

[6] Somewhat confusingly, sometimes the term 'satisfaction' is used here: $\mathfrak{M}$ is a model of Γ just in case $\mathfrak{M}$ satisfies Γ.

5 Kuhn: scientific revolutions as paradigm changes

5.1 Introduction

Our discussion of logical empiricism focused on efforts to articulate a 'scientific philosophy' centered upon the verification principle. Vienna Circle philosophers and scientists also pursued the project of *unified science*. To advance this project, the logical empiricists undertook to assemble the *International Encyclopedia of Unified Science*. As Otto Neurath, a social scientist, philosopher, and socialist reformer, wrote in the first volume of the *Encyclopedia*, "To further all kinds of scientific synthesis is one of the most important purposes of the unity of science movement, which is bringing together scientists in different fields and in different countries, as well as persons who have some interest in science or hope that science will help to ameliorate personal and social life" (Neurath, 1938/1955, 1).

In this chapter we will consider the philosophy of Thomas Kuhn, much of whose impact upon philosophy of science resulted from work that first appeared under the aegis of the unified science movement, yet whose views came to be seen as an emphatic rebuttal of at least some of the primary logical empiricist commitments of that movement.

Between 1938 and 1970, University of Chicago Press published the first two volumes of the *Encyclopedia* under the series title *Foundations of the Unity of Science*, with individual articles appearing as monographs. Volume II, number 2 appeared in 1962. Written by Kuhn and titled *The Structure of Scientific Revolutions* (Kuhn, 1962, hence, *SSR*),[1] the book came to be seen as contributing to the demise of the logical empiricist movement from which

[1] *SSR* went through three editions. The second edition (1970) added a lengthy postscript, but left the text otherwise nearly unaltered. The third edition (1996) added an index. All page citations here are to the third edition.

the *Encyclopedia of Unified Science* was born (but see Reisch, 1991 for a corrective to this received view). According to this view, the eclipse of the logical positivist project in philosophy of science by a 'post-positivist' philosophy of science was partially driven by a book published as part of the positivists' own encyclopedic undertaking.

Kuhn sought to understand scientific revolutions from the perspective of a *historical* study rather than through the use of logical analysis. In this way, Kuhn's project fundamentally differs from that of the logical empiricists. Some of Kuhn's apparent conflicts with the logical empiricists thus arise from their pursuit of different questions. Nonetheless, we will note some genuine points of disagreement.

5.2 Paradigms and normal science

Kuhn's historical studies convinced him that scientific knowledge develops through two quite distinct patterns or modes. In the first mode, which he called *normal science*, change accumulates slowly and research is "firmly based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice" (*SSR*, 10). Normal science, as Kuhn describes it, proceeds amidst a general agreement within a scientific community about what problems most need to be solved and what standards are to be used in the evaluation of efforts to solve those problems. Kuhn emphasized that such consensus is best understood as a by-product of the acceptance by the community of certain "past scientific achievements" as *exemplars* of good scientific work.

Kuhn called such exemplars *paradigms*, borrowing a term used to describe examples of grammatical forms studied in the learning of languages. But he used the word in other ways as well. The term *paradigm* took on a special importance in Kuhn's work and he used it throughout *SSR*, denoting sometimes a kind of theoretical framework, sometimes something more pervasive like a "worldview," sometimes a scientific community united by its commitment to such a worldview, sometimes a line of research aimed at a particular set of problems, and so on. One critic, citing Kuhn's "quasi-poetic style," counted "not less than twenty-one different senses" in which Kuhn used the term in *SSR* (Masterman, 1970, 61–65). In a postscript to the second edition of *SSR*, Kuhn sought to bring order to the thicket of

paradigm-locutions that appeared in the book by introducing a distinction between two primary senses of the term. First, there is the paradigm as the "entire constellation of beliefs, values, techniques," as well as generalizations and models that are shared by a scientific community, for which he introduced the term "disciplinary matrix" (*SSR*, 175). The second paradigm notion is the exemplar: "the concrete problem-solutions that students encounter from the start of their scientific education" and that "show them by example how their job is to be done" (*SSR*, 175).

The stability of a paradigm in both senses, Kuhn urged, is needed in order to sustain normal science. Kuhn emphasizes that we can expect to find paradigms wherever there is a group of researchers that regularly produces solutions to scientific *puzzles* and that regularly reaches fairly uniform consensus about whether particular efforts succeed in solving those puzzles. Such paradigms will vary greatly with regard to their size, longevity, and degree of specialization. Neo-Darwinian evolutionary biology is a sprawling discipline with many sub-fields, for example, while molecular phylogenetics constitutes a more narrowly focused (though still quite ambitious) endeavor.

5.3 Anomaly and crisis

No paradigm could be expected to be sufficiently rich and stable to serve as the basis of solutions to all of its problems. Thus, Kuhn argued, a second mode of scientific change comes in time to displace the gradual development of normal science: *revolutionary science*. Even an enterprise as successful and long-lasting as Newtonian celestial mechanics eventually encounters problems that cannot be solved within its own paradigmatic constraints. A problem that resists solution comes to be seen not simply as a problem of the ordinary sort but an *anomaly*.

Imre Lakatos once wrote that each scientific theory "at any stage of its development, has unsolved problems and undigested anomalies. All theories, in this sense are born refuted and die refuted" (Lakatos, 1978, 5). When a new theory is introduced some phenomena within its domain appear difficult to reconcile with that theory; these may never be made to agree satisfactorily with the theory. Kuhn agreed with Lakatos on this point. For Kuhn, the demise of any normal science always starts with some such anomaly.

One of Kuhn's own examples comes from the history of quantum physics. Physicists early in the twentieth century developed what came to be called

the *Old Quantum Theory*. The iconic achievement of Old Quantum Theory was Niels Bohr's 1913 model of the hydrogen atom. This was a *planetary* model of the hydrogen atom in the sense that it assumed a nucleus around which the electron travels in well-defined orbits. But the model allowed only certain orbits, corresponding to specific discrete energy levels for the atom as a whole. If the electron underwent a transition from a higher-energy state to a lower-energy state it would emit electromagnetic radiation carrying energy equal to the difference between the two, but it did so *without* passing through the continuum of energies between its initial and final state. In this way, Bohr sought to explain the fact that hydrogen gas emits light in a discrete spectrum.

Bohr, Arnold Sommerfeld, and others applied the Old Quantum Theory successfully to a number of problems, particularly those concerning spectra emitted by certain elements. Old Quantum Theory served well for calculating spectra for "two body" systems. Strictly speaking, that means just hydrogen (one proton, one electron), but the same treatment was successfully extended to 'hydrogenic' atoms that approximate a two body system. In such hydrogenic atoms, there is an inner core of electrons that are closely bound to the nucleus, such that they together with the nucleus can be treated as one body, with a single orbiting electron in addition. (Sodium is an example.) But the theory did not work for three body systems. It also made incorrect predictions regarding the splitting of spectral lines in a magnetic field, known as the *Zeeman effect*.[2] Such problems led to a situation that Kuhn characterizes as a *crisis* for the Old Quantum Theory.

A Kuhnian *crisis* occurs when members of a research community find that some anomaly (or set of anomalies) has proven so resistant to solution within the current paradigm that they begin to despair of thus resolving it. Since it is an imperative of every paradigm to solve all of the anomalies in its domain, members of the community increasingly feel that they are called upon to do the impossible. Kuhn emphasized the psychological dimension of such crises, which he documents with a quotation from Wolfgang Pauli, writing to the young physicist Ralph Kronig during the crisis of the Old Quantum Theory: "At the moment physics is again terribly confused. In any case, it's

[2] More precisely, the Old Quantum Theory predicted a splitting into triplets, an observed phenomenon known as the 'normal Zeeman effect.' But in the *anomalous Zeeman effect*, spectral lines split into four or six components in a magnetic field.

too difficult for me, and I wish I had been a movie comedian[3] or something of the sort and had never heard of physics" (*SSR*, 84, quoting from (Kronig, 1960)).

5.4 Revolution

Driven by growing doubts about the ability of the paradigm to resolve its own anomalies, members of the community begin to consider routes to a solution that are at odds with the paradigm's own constraints. Thus begin revolutions in the sciences, as Kuhn depicts them. In the case of Quantum Theory, Werner Heisenberg made the first breakthrough when, having spent time with Bohr in Copenhagen, he returned in April 1925 to his position in Göttingen to work by himself. The paper that he then wrote laid the groundwork for a radically new way of doing quantum mechanics, making use of a new mathematical apparatus and eliminating Bohr's planetary orbits. Heisenberg's theory predicted that quantum particles like electrons would have well-defined positions or momenta when those properties were measured, but did not describe determinate trajectories for quantum particles at other times. Heisenberg's paper (Heisenberg, 1925), together with subsequent joint efforts involving Max Born and Pascual Jordan (Born & Jordan, 1925; Born, Heisenberg, & Jordan, 1926), laid the foundations for the New Quantum Theory.

Heisenberg's proposal was both exciting and controversial. Motivated by a roughly positivist idea (but see Bokulich 2008), Heisenberg saw Bohr's 'unobservable' atomic orbits as a theoretical defect. Albert Einstein objected to this reasoning and regarded the theory that resulted from Heisenberg's efforts at removing this defect as an incomplete account of mechanics at the atomic scale. On Kuhn's account, this is typical of revolutions. The community does not convert as a whole all at once. Some will grasp immediately for the relief from anomaly offered by a new proposal. Kuhn quotes Pauli, writing again to Kronig, shortly after the appearance of Heisenberg's 1925 paper: "Heisenberg's type of mechanics has again given me hope and joy in life. To be sure it does not supply the solution to the riddle, but I believe it is again possible to march forward" (*SSR*, 84, quoting from Kronig (1960)).

[3] Comments that he resembled the film actor Edward G. Robinson delighted Pauli (Enz, 2002).

In a revolution, scientists accept both a bold new theory and a new way of doing science, including new ways of deciding what constitutes an adequate solution to a problem. Recall that central to the idea of a paradigm is the acceptance of an exemplar that serves as a model of good work in the discipline. These exemplars serve as the basis for the articulation of standards. Correspondingly, a revolutionary change involves the rejection of the old exemplars *and the accompanying standards of adequacy* by which problem solutions are to be evaluated. A candidate for a new paradigm must come with its own potentially exemplary solutions. In cases of revolution these will inevitably be found lacking by the old standards, but will succeed (inevitably) by the new standards (which after all are based on these new candidates).

The result is one of Kuhn's most controversial theses: During periods of revolution, choosing between paradigms cannot be guided by *accepted* standards at all. The choice between the old paradigm and a potential new one "is not and cannot be determined merely by the evaluative procedures characteristic of normal science, for these depend in part upon a particular paradigm, and that paradigm is at issue. When paradigms enter, as they must, into a debate about paradigm choice, their role is necessarily circular. Each group uses its own paradigm to argue in that paradigm's defense" (*SSR*, 94).

Many initial responses to *SSR* harshly criticized such claims, which appeared to depict scientific revolutions as events in which scientists changed theories without relying on anything resembling experimental reasoning or logical inference. Such an account seemed radically at odds with long-held views about the responsible conduct of science. If "[e]ach group uses its own paradigm" to defend its commitment to that paradigm, then those arguments can only appeal to those who already accept the paradigm in question. Acceptance comes first, arguments come later. Arguments over paradigm choice then seem more like armed conflict, or perhaps political campaigns, than rational deliberation on the basis of empirical evidence, since the arguments cannot do any work until one has somehow gotten one's opponent to adopt one's own point of view. And then, what work is left for the argument to do? As Dudley Shapere characterized the problem in an early and influential critical review of *SSR*, Kuhn's account entails that "the decision of a scientific group to adopt a new paradigm is not based on good reasons; on the contrary, what counts as a good reason is determined by the decision" (Shapere, 1964, 392).

Kuhn's own language and choice of analogies for explaining his view exacerbated the perception that his account of scientific revolutions is radically subjective. He compares a scientific revolution to a political revolution: The latter occurs when crisis becomes so acute that "political recourse fails"; likewise the former occurs in a crisis by methods going beyond the "evaluative procedure characteristic of normal science" (*SSR*, 93–94). He compares changes in paradigm to visual "gestalt shifts": "What were ducks in the scientist's world before the revolution are rabbits afterwards."[4] He uses language to describe changes in paradigm that highlight similarities with religious conversion: "The transfer of allegiance from paradigm to paradigm is a conversion experience that cannot be forced" (*SSR*, 151), and "[a] decision of that kind can only be made on faith" (*SSR*, 158).

5.5 Incommensurability

Crucial to Kuhn's depiction of the wide and multi-faceted separation between paradigms is the concept of *incommensurability*. Literally, incommensurability means 'lacking a common measure,' a phrase indicative of the problem faced by scientists in a period of Kuhnian revolution: one needs to choose a paradigm, but lacks a common measure for determining which of the competing candidates is *better*.

A careful reading of *SSR* reveals an important distinction with regard to incommensurability. Rival paradigms are incommensurable in both a *methodological* and a *semantic* sense (Bird, 2000). Let us consider methodological incommensurability first.

Suppose that in a period of crisis for paradigm P_1 a rival paradigm P_2 takes shape, such that members of the community must make a choice to continue working within P_1 or switch their allegiance to P_2. One might suppose that this would be a matter of evaluating the extent to which the theoretical content of each of the paradigms is supported by the available data. According to the thesis of methodological incommensurability, however, the methods and standards by which such support are to be determined are partially

[4] As Kuhn himself noted, N. R. Hanson had gone much further in the discussion of gestalt shift images as elucidating some similar views about how theoretical commitments affect one's empirical evidence (Hanson, 1958).

determined by the very paradigms under consideration: "the choice is not and cannot be determined merely by the evaluative procedures characteristic of normal science, for these depend in part upon a particular paradigm, and that paradigm is at issue" (*SSR*, 94).

Semantic incommensurability adds to the difficulty. According to Kuhn, terms mays be used in competing paradigms so differently that they have different meanings. Kuhn's example in *SSR* involves the term 'mass' as used in Newton's dynamics and in Einstein's dynamics. For Newton, mass is an invariant quantity, i.e., the mass of a given body has the same value regardless of the state of motion of that body relative to a given observer. In Einstein's theory, however, two observers who are in motion relative to one another will disagree in their measurements of the mass of a body (though the difference will be negligible if the relative velocities are small compared with the speed of light). Kuhn argued that such an important difference entails that 'mass' does not mean the same thing in Newton's physics as it does in Einstein's.

Semantic incommensurability points to perhaps the clearest disagreement between Kuhn's account and logical empiricism. A logical empiricist interpretation of this state of affairs would insist upon the distinction between the theoretical and observational vocabulary of a theory. Since Newton's dynamical theory and Einstein's dynamical theory say such different things about mass, this must be regarded as a theoretical term, not an observational one. Each theory would then use distinct bridge principles to connect statements involving mass to statements in the observational vocabulary, and the choice between theories would be based on seeing which theory entailed observation-vocabulary statements in better agreement with the results of actual observations.

This reply assumes that we can draw a principled and well-defined distinction between observational and theoretical vocabularies – in short, that observation vocabulary is *theory neutral*. Kuhn, however, rejects any such distinction. He not only rejects the idea of a theory-free observational vocabulary, he suggests that in some sense *there is not even a paradigm-independent world to be observed*: "though the world does not change with a change of paradigm, the scientist afterward works in a different world" (*SSR*, 121).

Kuhn draws this conclusion from the ways in which scientists are trained to make observations of a certain kind, and incorporate apparently theoretical concepts and languages into their observations, thus making observation *theory-laden*: "Looking at a contour map, the student sees lines on paper, the

cartographer a picture of a terrain. Looking at a bubble-chamber photograph the student sees confused and broken lines, the physicist a record of familiar subnuclear events. Only after a number of such transformations of vision does the student become an inhabitant of the scientist's world, seeing what the scientist sees and responding as the scientist does" (*SSR*, 111). But this training takes place *within* a paradigm (a disciplinary matrix), so that when a revolution occurs "the scientist's perception of his environment must be re-educated – in some familiar situations he must learn to see a new gestalt" (*SSR*, 112). How precisely to understand such passages is not perfectly clear, but they convey at least this much: one cannot hope to resolve disputes over rival paradigms by adopting a description of the relevant data that is free of assumptions belonging to the very paradigms at issue.

5.6 Re-articulation

Regardless of Kuhn's intent, many readers of *SSR* took him to be "attempting to undermine the cognitive authority of science rather than to suggest a different view of its nature," as Kuhn himself later acknowledged (Kuhn, 1993, 314). To counter this reading of his work, Kuhn emphasized in later publications that some continuities do persist through even the most revolutionary changes in science. Scientists operating in any paradigm will, for example, value accuracy in a theory's predictions, consistency in its central assumptions, simplicity in its formulation, breadth in its scope, and fruitfulness in its applications. Such continuities, however, do not suffice to secure a neutral standpoint from which to evaluate paradigms because these *shared values underdetermine theory choice*.

Kuhn's underdetermination thesis rests on two observations. First, each of these values is, on Kuhn's account, *imprecise* or *ambiguous* in their application. Just what it means for a theory to be simple is, for example, notoriously difficult to define rigorously, and scientists have often disagreed in their judgments of simplicity. Second, scientists may disagree about the *relative importance* of these values, a problem aggravated by the fact that some of them are in tension with others. One might, for example, be able to make one's theory more accurate by making it less simple. The problems of imprecision and differential relative weighting of values both entail, according to Kuhn, that different scientists may reasonably make different choices among competing paradigms (Kuhn, 1977). Indeed, according to Kuhn, the historical

progression of science shows that "before the group accepts it, a new theory has been tested over time by the research of a number of men, some working within it, others within its traditional rival," and this "mode of development … *requires* a decision process which permits rational men to disagree" (Kuhn, 1977, 332; emphasis in original).

When critics such as Shapere allege that according to Kuhn paradigm choice "cannot be based on good reasons of any kind" (Shapere, 1966, 67), Kuhn argues, they in effect accuse him of "throwing out the baby with the bathwater." Kuhn's re-articulation seeks to clarify what he regards as the baby to be retained (that scientific judgments, even in times of revolution, involve reasoned judgments) as opposed to bathwater to be discarded (that such judgments are guided by "algorithmic" rules that when correctly applied will lead all competent and rational scientists to the same conclusions). Paradigm choice may not be determined by any uniform and unequivocal set of rules, but this does not mean that it is not based on good reasons (or "good sense"). Good reasons functioning within incommensurable paradigms, however, may yield conflicting results.

5.6.1 An example: molecular phylogenetics

At this point, it might help to have an example. Let us briefly view, through a Kuhnian lens, the field of molecular phylogenetics. We will find that applying Kuhn's ideas to a given scientific discipline is not an entirely straightforward matter.

Phylogenetics constitutes the approach that contemporary evolutionary biologists take to *systematics*, which is the study of relationships among biological taxa. Phylogeneticists attempt to infer from various kinds of data how closely related different groups of organisms are. That is, they seek to determine, among a given group of distinct populations (such as species), which groups share a more recent, and which a more remote, common ancestor. The results of such inferences are represented in the form of *phylogenetic trees*, showing the branching events at which different groups of organisms became hereditarily distinct (see Figure 5.1). To make reliable inferences about phylogeny, one must first determine what observations are relevant for determining degrees of relatedness among taxa. It would be natural to suppose that groups that share more traits must share a more recent common ancestor than groups that share fewer traits, were one to

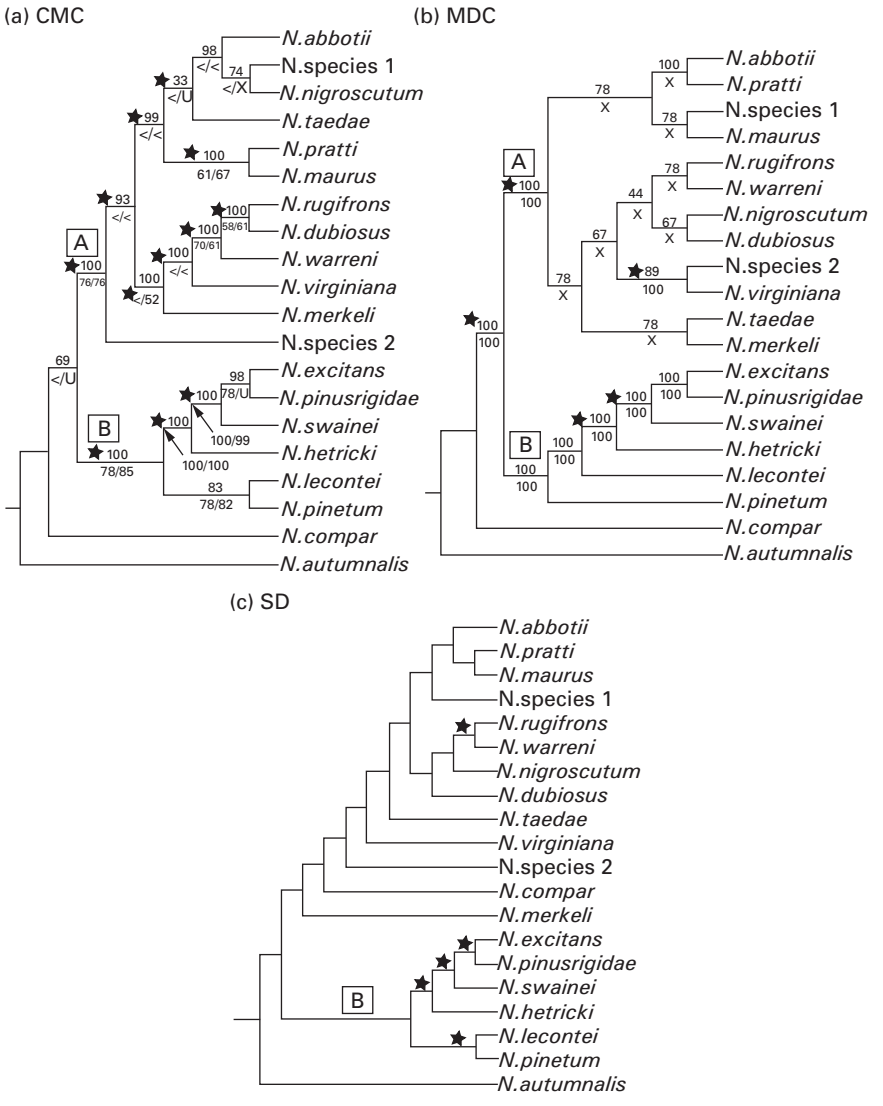


Figure 5.1 Three phylogenetic trees of the sawfly (genus *Neodiprion*). The different trees result from different methods applied to a single data set. Although seemingly a normal science, phylogenetics, like many disciplines, sustains continued debates over the relative merits of competing methods (Linnen & Farrell, 2008; by permission of Oxford University Press).

ignore the fact that traits can be shared without having been inherited from a common ancestor. In any case, such a principle is useless unless one knows exactly what to count as a trait and when to consider a trait to be shared among two taxa. A further question concerns how much

significance to attribute to different kinds of traits as indicators of common descent.[5]

With the availability of a number of powerful experimental techniques that allow for the generation of large data sets, *molecular phylogenetics* has transformed contemporary phylogenetics, using genetic data to arrive at phylogenetic conclusions. Earlier systematicists based their conclusions on observations of the form and structure (*morphology*) of organisms. But phylogenetics itself evolved with the emergence of methods of genome sequencing, facilitated by methods for producing larger numbers of copies of strands of DNA, thus *amplifying* the genetic information available in a sample. Molecular phylogenetics, thus, looks like the outcome of a Kuhnian revolution within phylogenetics that has introduced new laboratory methods and training, changed the kind of data regarded as most important, and introduced new problems requiring solution.

Suppose, for example, that one is interested in determining just *when* the ancestry of two distinct contemporary taxa parted ways. Molecular phylogeneticists can measure a kind of 'genetic distance' between two organisms, but to translate that into an estimate of how long ago their ancestors diverged, one needs a good estimate of the rate at which molecular sequences change. However, the rates of such divergence are highly variable, complicating the task of reliably estimating times of divergence. Morphologically based phylogenetics did not face this problem; as a rule, in that tradition one based estimates of times of divergence on the fossil record, when fossils were available that were relevant. If such fossils were rare, at least they could be dated fairly reliably. By contrast, DNA is readily available from living organisms, greatly facilitating comparisons of genetic resemblance when fossil evidence is lacking.

Molecular phylogenetics seems to have its own exemplars as well, which, though they share some features with earlier morphological exemplars, are distinguished by the kind of data they use and methods of analysis that are applied. Such exemplars, as Kuhn emphasized, are especially prominent in textbooks that are used in the training of members of the community. However, consulting a textbook such as *Molecular systematics* (Hillis, Moritz, & Mable, 1996), one finds many exemplars from earlier morphological

[5] That the solution to such problems by phylogeneticists is a matter of "persuasion, background, training and tradition" just as Kuhn suggests has been argued by Richards (2002).

approaches, since many of the methods of analysis are the same (such as the use of parsimony considerations in selecting from among the various possible phylogenies that can be generated from a given set of data). We might take such co-existence of exemplars from two distinct approaches either as an indication that molecular phylogenetics is not after all such a revolutionary (in Kuhn's sense) development in systematics, or as an indication that Kuhnian revolutions need not result in the complete replacement of the traditions and exemplars of the earlier paradigm.

5.7 Progress through revolutions?

The final chapter of *SSR* speaks significantly to Kuhn's own assessment of the real novelty of his account. He invokes an analogy with another revolutionary idea in science: Darwinian evolution. What most shocked Darwin's contemporaries, according to Kuhn, was "neither the notion of species change nor the possible descent of man from apes," since evolutionary ideas had been in circulation for some time. What really distinguished Darwin's theory of evolution from the ideas developed by others was that "[t]he *Origin of Species* recognized no goal set either by God or nature." Nonetheless it described a process of "gradual but steady emergence of more elaborate, further articulated, and vastly more specialized organisms" a process that Darwin himself understood as one of advancement or progress (*SSR*, 171–172).

Likewise, Kuhn suggests, we should regard the development of science as progressive: "The net result of a sequence of such revolutionary selections, separated by periods of normal research, is the wonderfully adapted set of instruments we call modern scientific knowledge," and the process is "marked by an increase in articulation and specialization." Yet the process need not be understood as directed toward "a set goal, a permanent fixed scientific truth" (*SSR*, 172–173). Science does progress, according to Kuhn, but the incommensurability of earlier and later paradigms poses an obstacle to the idea that there is a "coherent direction of ontological development" in a revolutionary transition. On the contrary, "the notion of a match between the ontology of a theory and its 'real' counterpart in nature now seems to me illusive in principle." Kuhn's view turns out then to be a version of *anti-realism* about science (*SSR*, 206; McMullin, 1993).

We will take up general debates between realists and anti-realists in Chapter 10. For now, we might wish to note a tension between Kuhn's

rejection of science as aimed at truth and his embrace of the idea that science progresses. If science does not progress toward truth, and is marked by transitions between incommensurable paradigms, how can we meaningfully consider it progressive? According to a recent study by K. Brad Wray, Kuhn's notion of progress is precisely that of increasing "articulation and specialization." Wray argues that the increasing specialization of science should be understood not merely as a sociological phenomenon but as serving important epistemic functions, enabling scientists to enhance the accuracy of their work by focusing on ever-narrower domains (Wray, 2011).

5.8 Problems with paradigms

Arguments against the central tenets of Kuhn's account have flourished in proportion to the attention his work has received. Unable to adequately survey the range of issues that have been raised in a brief discussion, I will mention just two.

Incommensurability would seem to rule out the possibility of a paradigm-neutral evaluation of the evidential significance of a given observation, since the terms in which evidence is described and evaluated will take on different meanings in incommensurable paradigms. Historically, however, the evidential significance of particular findings very often *survives* revolutionary changes in theory.

Consider an example from the very beginnings of evolutionary phylogenetics. In *Origin of species*, Charles Darwin noted the structural commonalities between a bat's wing and (for example) the hands of humans as evidence that these distinct species share a common ancestry (Darwin, 1859, 434). Darwin was able to make this inference in spite of the fact that he had no good theory of the physical basis of inheritance. A revolution in biology brought together Darwin's theory of evolution by natural selection and the genetic theory of inheritance that emerged from work done by Gregor Mendel in the nineteenth century and developed by an extended experimental and theoretical undertaking during the first two decades of the twentieth century. After this revolution in our understanding of genetics and of evolution, that same shared morphology between bats and primates continued to serve as evidence of common ancestry, although just what that involves had been given a much deeper level of explanation.

Moreover, there is good reason to think that scientists at least sometimes *anticipate* significant changes in theorizing (without knowing just what those changes will be) and take measures in their experimental work to secure their evidence against changes in theories. For example, when he was performing experiments that provided crucial insights into the nature of electromagnetism in the middle of the nineteenth century, Michael Faraday took careful measures to describe his experimental results in a manner that would make them independent of any of the theories of electricity or magnetism of his day (Cobb, 2009). Isaac Newton, in the course of arguing for his Universal Law of Gravitation, ensured that his inference, though extremely bold in its scope, was free of any commitment to particular views about the cause of gravity, an issue that he regarded as unsettled (Harper, 2002, 2011).

Such efforts need not presume the availability of a theory-neutral observation language of the sort espoused by logical empiricists. Mary Hesse, for example, combines an embrace of Kuhn's thesis of theory-ladenness of observation with a recognition of the possibility of paradigm-neutral theory testing in her *network model* of scientific theories (Hesse, 1980). Hesse's argument relies on the observation that even where a term used in reports of observations does have different meanings in different theoretical networks, there will often be an area of "*intersection* of predicates and laws between the theories," the exploitation of which can yield paradigm-independent judgments of the viability of competing theories (Hesse, 1980, 97).

Another problem that Kuhn's account faces concerns his claim that scientific revolutions involve changes in methodological standards of appraisal. This claim is crucial to his assertion that "When paradigms enter, as they must, into a debate about paradigm choice, their role is necessarily circular" (*SSR*, 94). Philosophers of science have proposed numerous accounts of scientific method that aim at a kind of generality that would transcend commitments to particular theories at particular times; if Kuhn is right, they are on a wild goose chase. But whether or not any particular one of these accounts is entirely satisfactory, there is some reason to think that scientific inquiry relies, to a very large extent, on methods of appraisal that at least survive changes in particular theories in particular disciplines. Moreover, when such methods do change, as they sometimes do, it is not because of a change in any particular theory to which those methods might be applied.

A good example would be the history of developments in theoretical statistics that have provided scientists in essentially every discipline with

new tools for evaluating the reliability and probative power of their testing procedures, and for making the most efficient use of the data they produce. These developments have a fascinating history of their own (Hacking, 1990; Porter, 1988; Stigler, 1990). Even if a Kuhnian view of the history of statistics would regard these developments as involving one or more revolutions, the methods developed out of the advances in statistical theory provide an ongoing basis for building up and evaluating knowledge from experimental data that does not depend on the reigning paradigm in any particular discipline, whether it is epidemiology or cognitive neuroscience or astrophysics.

5.9 Conclusion

Scholars continue to debate the significance of Thomas Kuhn's work and the model of scientific development that is articulated in his writings (Fuller, 2001; Hoyningen-Huene, 1993). He played a significant role in an important transformation of the discipline. He combined an intriguing perspective on scientific change with a distinctive kind of reasoning that really was different from his predecessors. His historical approach gave priority to the study of episodes of scientific controversy and discovery from the perspective of the participants in those episodes. At least in principle, he intended to let the study of such episodes come first, and the philosophical lessons to be drawn come second. This priority is reflected in a rhetorical question posed in the final chapter of *SSR* about judgments of whether a problem has been adequately solved: "What better criterion than the decision of the scientific group could there be?" (*SSR*, 170). The question reflects Kuhn's reticence about philosophical attempts to formulate independent criteria of rationality or progress or adequacy to which scientists might then be held accountable.

This is not to say that Kuhn held that *anything goes* in science. Not just any set of ideas and methods can sustain a normal science tradition. In that sense, Kuhn regarded his account as giving some idea of what a community of inquirers *must be like* in order to constitute a scientific community. On the other hand, a functioning normal science tradition is, in a certain sense, beyond reproach on this account. More precisely, any criticism can only find a legitimate footing within the tradition itself, at least until a crisis comes along.

6 Lakatos: scientific research programs

6.1 Introduction

In this chapter and the next we will consider the contributions of Imre Lakatos and Paul Feyerabend, respectively. They were contemporaries of Thomas Kuhn who shared some of Kuhn's commitments but developed them into strikingly different visions of the scientific enterprise.

Like Kuhn, both Lakatos and Feyerabend emphasize that an adequate philosophy of science has to answer to the history of science (though they disagreed on the exact role history should play). Like Kuhn, they insist that the theoretical framework in which a scientist works influences the way that she interprets her evidence, making the logical empiricists' sharp distinction between theoretical and observational vocabularies untenable.

Yet Kuhn, Lakatos, and Feyerabend offer very distinct philosophies with different aims. An important difference concerns the *rationality* of science. For Kuhn, an adequate account of revolutionary science must allow for rational disagreement among scientists, but beyond that rationality is not central to his analysis. For both Lakatos and Feyerabend, the problem of rationality *is* central, but they take positions that are, at least on the surface, diametrically opposed. Lakatos seeks to *demonstrate* the (at least partial) rationality of the historical progression of scientific thought (and to distinguish it from irrational pseudoscience). Feyerabend, on the other hand, argues that any attempt to constrain science by a philosophical model of rationality (whether that of the logical empiricists, or Popper, or anyone else) will prove counter productive, inhibiting the free development of the individual scientist and blocking the growth of scientific knowledge. Both disagreed vehemently with what they regarded as Kuhn's deference to the consensus of the 'normal science' community. Perhaps not coincidentally, Karl Popper was influential in the development of both Lakatos's and Feyerabend's philosophies of science.

The present chapter will seek to highlight the main features of Lakatos's account of *progressive research programs*.

6.2 Falsification and the underdetermination problem

Lakatos's life story deserves a longer telling than I can give here.[1] He was born into a Jewish family in Hungary in 1922 (his family name was Lipsitz). When Hungary entered World War II on the Axis side, he adopted a false identity, first to avoid being forced into a 'labor company' by the Hungarian government, then to escape deportation to an extermination camp when the Nazis occupied Hungary in 1944 (his mother and grandmother were killed in Auschwitz). He adopted the name Lakatos after the Russians invaded Hungary. After studying mathematics, physics, and philosophy[2] he held a post in the post-war Hungarian Ministry of Education, but was then imprisoned for nearly four years on charges of 'revisionism.' When anti-Soviet resistance swelled up in 1956, Lakatos helped write a statement proclaiming freedom of scientific inquiry on behalf of the National Committee of The Hungarian Academy of Sciences. Shortly after Soviet tanks rolled into Budapest to put down the uprising, Lakatos emigrated, first to Vienna, then to Cambridge, where he wrote a PhD thesis in philosophy of mathematics. He began an appointment at the London School of Economics (LSE), joining Karl Popper in the Department of Philosophy, Logic and Scientific Method that Popper had founded. He was appointed Professor of Logic in 1969, but died in 1974 at the age of 51.

He began to develop his views about the philosophy of the natural sciences after his appointment at LSE, and he did not live to write the book that would fully articulate his views. Nonetheless, he wrote a long essay titled "Falsification and the methodology of scientific research programmes," first published in a collection of essays arising from a symposium held in 1965 in which both Kuhn and Popper participated (Lakatos & Musgrave, 1970). The differences between Kuhn's and Popper's philosophies are central to the

[1] See (Larvor, 1998) for a detailed account of Lakatos's life, as well as an insightful study of his philosophy.

[2] A comprehensive understanding of Lakatos's philosophy would require attention to his study of the philosophy of Hegel (Larvor, 1998; Kadvany, 2001).

papers collected in that volume, and constitute a significant focus of Lakatos's paper as well.

Lakatos states explicitly what is at stake in debates over scientific methodology. In his introduction (titled "Science: Reason or Religion?") he states: "The clash between Popper and Kuhn is not about a mere technical point in epistemology. It concerns our central intellectual values, and has implications not only for theoretical physics but also for the underdeveloped social sciences and even for moral and political philosophy." He continues with a polemic that defenders of Kuhn would regard as distinctly uncharitable: "If even in science there is no other way of judging a theory but by assessing the number, faith and vocal energy of its supporters, then this must be even more so in the social sciences: truth lies in power. Thus Kuhn's position vindicates, no doubt, unintentionally, the basic political *credo* of contemporary religious maniacs ('student revolutionaries')" (Lakatos, 1970, 93).

Lakatos's account of scientific change seeks to demonstrate how a broadly falsificationist philosophy of science can deal adequately and explicitly with the problem of the underdetermination of decisions regarding theories by observations alone (including decisions to regard them as falsified). Lakatos regarded Popper's falsificationism as a failure in this regard.

Lakatos begins his critique of Popper's falsificationism by distinguishing it from a distinct position with which Popper's is sometimes confused. The latter he refers to as "dogmatic falsificationism," while he considers Popper to have advocated "methodological falsificationism." The distinction concerns the status of statements describing empirical results that might serve to falsify theoretical statements. For the dogmatic falsificationist, these statements are simply those that report the results of *observation*, which is to be regarded as a distinct "psychological" process from *theorizing*. An observational statement, according to this view, is *proved* (or at least justified) by the observation of facts.

The methodological falsificationist (such as Popper himself) rejects this untenable claim. As already discussed in Chapter 2, Popper denied that experiences such as the observation of a fact could justify statements because the only thing that could do that would be another *statement* that logically entailed the first. Neither psychological states ("in the mind") nor states of affairs ("out in the world") can play this role. Contrary to the views of the logical empiricists, statements used to describe experimental or observational results, and potentially to falsify theories, are articulated in language already

bearing theoretical commitments. Their acceptance as part of the empirical basis results from the *decision* of the investigator.

But Lakatos emphasizes an important difference between *naïve* and *sophisticated* methodological falsificationism. For naïve methodological falsificationists (NMFs), a theory is "'acceptable' or 'scientific'" if it is falsifiable by experiment. The sophisticated methodological falsificationist (SMF) considers a theory to acquire such status only if it "leads to the discovery of novel facts" (Lakatos, 1970, 116). The NMF considers a theory to be falsified when he decides that a statement accepted as part of the empirical basis conflicts with that theory. For SMFs, a theory T is falsified only if some *other* unrefuted theory T' has been proposed that predicts *novel* facts (facts that T predicts will not hold, or at least that it presents as very improbable), at least some of which have been corroborated, and that explains facts previously explained or predicted by T.

6.3 The methodology of scientific research programs

In the progression from dogmatic falsificationism to sophisticated methodological falsificationism Lakatos's focus shifts from the question of whether an individual theory is scientific to the question of what, in his words, "intellectual honesty" requires of investigators by way of response to a potential falsification of a given theory. Lakatos proposes a parallel shift from judging theories to judging *sequences of theories* that bear a certain historical relationship to one another. These sequences of theories he calls *scientific research programs*, and the centrality of their role in his methodology marks Lakatos's most significant departure from Popperian falsificationism.

More precisely, a scientific research program is a historical entity with four main components. The program must include (i) a single *hard core* of "assumptions which, by methodological decision, as it were, are kept unfalsified" (Zahar, 1973, 100). To obtain a theory, one must conjoin the assumptions from the hard core with (ii) auxiliary assumptions constituting the *protective belt*; these assumptions are expendable, and when a potential falsification or anomaly arises for a theory thus constructed within the research program, investigators direct falsification at the protective belt, rather than the hard core. Methodologically, the construction and subsequent modification of a scientific research program is guided by (iii) a *negative heuristic*, which directs the investigator to deflect potential falsifications away from the hard core and

toward statements in the protective belt, and (iv) a *positive heuristic*, a kind of guideline or policy for the initial construction and subsequent modification of the protective belt. Thus, a research program comprises a series of theories (a 'problemshift', in Lakatos's terminology) as investigators combine the unchanging hard core with the changing contents of the protective belt.

Let us consider an example from the history of physics, of the sort Lakatos had in mind in developing his methodology, to see how we might apply these ideas.

6.3.1 A tale of two planets

The planet Uranus was discovered by William Herschel in 1781, but over the following years observations of Uranus proved difficult to reconcile with predictions drawn from Newtonian Celestial Mechanics (NCM: essentially, this is Newton's laws of motion and law of gravity, mathematically elaborated through the subsequent century and applied to the known bodies of the solar system). In 1845, Françoise Arago, who was then director of the Paris Observatory, encouraged a young astronomer–mathematician named Urbain Le Verrier to work on the anomalous observations of Uranus's orbit.

Other astronomers had speculated that the problem might arise from the gravitational field of "a new planet whose elements would be known according to its action on Uranus" (Wilhelm Bessel, writing in 1840, quoted in Hanson, 1962, 361). Le Verrier set to work to calculate, using NCM: (i) the precise effect of Jupiter and Saturn on Uranus's orbit; (ii) the exact disagreement between the predictions and the observations of Uranus's orbit, accounting for the perturbations determined in his first step; and (iii) what hypothetical properties of a possible new planet would introduce a further perturbation in Uranus's orbit sufficient to reconcile prediction and observation.

Le Verrier presented his work in Paris, but did not succeed in persuading anyone there to search for the planet that his analysis predicted, so he wrote to Johann Gottfried Galle at the Berlin observatory. Galle's student, Heinrich Louis d'Arrest, suggested making a comparison with a recently constructed chart of the stars and, just after midnight on September 24, 1846, the Berlin astronomers found the planet. It lay within 1° of Le Verrier's predicted location and had an apparent diameter in very close agreement with the predicted value.

Independently of Le Verrier, John Couch Adams had taken up the problem tackled in part (iii) of Le Verrier's calculations. He had even less success

than Le Verrier in persuading others to take his calculations seriously. He communicated his results to James Challis, director of the Cambridge Observatory, and to George Airy, the Astronomer Royal.[3] Only when Airy got wind of Le Verrier's Paris presentations did he set Challis's observatory to work on a search for the planet, which began on July 29, 1846. Challis did not have as good a star map as Galle, and only after the announcement of the Berlin observation did the English astronomers realize that they had observed the planet in August without realizing it.[4]

Note that by the criteria of NMF, these astronomers made the wrong move. When they 'observed' the conflict between the data on Uranus's orbit and the prediction from NCM, they should have *abandoned* Newton's theory as falsified. Instead, they proposed to save the theory by adopting a new auxiliary assumption. SMF does not explicitly condemn this response, but does not tell us why it was reasonable to introduce a new auxiliary hypothesis *rather than* trying to come up with a replacement for the core commitments of NCM.

Lakatos's methodology seeks to improve upon NMF and SMF by doing two things: First, it articulates a new unit of methodological appraisal (the scientific research program) and its components (hard core, protective belt, and negative and positive heuristics). Second, it articulates criteria by which to *evaluate* the responses of scientific research programs to potential falsifications.

Applying Lakatos's analysis of research programs to our example, Le Verrier and Adams regarded NCM as the hard core. The anomalous features of Uranus's orbit presented a problem to be solved, but, in keeping with the negative heuristic, not a reason to reject NCM. Furthermore, they regarded statements about the bodies to be found in the solar system as lying within the protective belt. That belt had been modified before (when Uranus was added to the solar system), and could be modified again. Plausibly, the positive heuristic they followed states that one may hypothesize new bodies with

[3] One aspect of the multi-faceted controversy surrounding this episode concerns why neither Airy nor Challis saw fit to follow up on Adams's prediction (see Sheehan, Kollerstrom, & Waff, 2004; Standage, 2000 for two engaging accounts).

[4] Accounts of the episode sometimes assert that Le Verrier and Adams independently "made the same calculations." In fact, Adams only tackled one part of Le Verrier's three-part project. Their resulting calculations disagreed with each other and with the actual orbit of Neptune, but all three happened to coincide for a brief period, during which the observations in Berlin and Cambridge were made (Hanson, 1962).

orbital parameters allowing their observation in order to explain otherwise anomalous orbits of celestial bodies.

6.4 Criteria for progress

Lakatos considered falsificationists to be concerned with determining whether a given theory is acceptable or scientific. A central motivation for the methodology of scientific research programs (MSRP) is that, in a certain sense, every theory is "born falsified" (in the sense that every theory has unresolved anomalies throughout its historical development), and yet theoretical principles can be rendered unfalsifiable by simply redirecting the inference that an error has occurred onto some statement in the protective belt. Moreover, sometimes it is reasonable to do this, as astronomers did with regard to NCM in the period 1781–1846.

Just as clearly, though, it is not *always* reasonable to protect a theory from falsification in this way. Sometimes the lesson to draw from an apparent failure of a theory is that the theory is wrong! We would like a principle that allows us to differentiate these two possibilities. Another episode from the history of astronomy illustrates why.

6.4.1 A tale of another planet?

Urbain Le Verrier, the hero of the discovery of Neptune, also worked on the orbit of Mercury. In the 1840s, he set out to determine the theoretical orbit of Mercury, for which very precise observational data were available. But he could not make the data fit the theory. Norwood Russell Hanson comments that for Le Verrier "the theory of Mercury was like a frayed garden hose conveying high-pressure steam. When he adjusted it here, it sprang a leak there; he hadn't enough hands, or ideas, to plug every hole at once" (Hanson, 1962, 366). In an 1849 report to the Paris Academy, he traced the difficulties to a mis-match regarding the precession of Mercury's perihelion (the point in the orbit at which the planet comes closest to the Sun). According to NCM, the perihelion of Mercury's orbit should advance 527″ of arc over the course of a century, but the observational data indicated an additional 38″ of advance. (In 1882 Simon Newcomb corrected the magnitude of the anomaly to approximately 43″/century.)

Le Verrier attempted in 1859 to reconcile theory and observation by adjusting the planet's orbital parameters (such as its mass), but these were already very narrowly specified by existing data and theory. If he changed them by very much, it ruined the fit between theory and data elsewhere. So he turned to the trick that had worked so well in the case of Uranus: he considered the postulate of a new planet (or "group of planetoids") inside Mercury's orbit, previously unobserved because of its (or their) proximity to the Sun. A new telescopic search was needed.

Many observational reports of an intra-Mercurial planet followed. (These described the observation of *transits*, in which a dark spot appeared in front of the Sun; such spots had to exhibit planetary motion, lest they turn out to be sunspots.) Le Verrier received one report in December 1859 from a provincial doctor and amateur astronomer named Edmond Modeste Lescarbault. Lescarbault's report was sufficiently credible that Le Verrier, by now the Director of the Paris Observatory and an important scientific dignitary, traveled to the town of Orgères-en-bouce to query Lescarbault directly. The doctor was convincing. Le Verrier announced the finding to the Académie des Sciences in January 1860 and arranged for Lescarbault to receive the Legion of Honor. Thus was the planet that came to be known as *Vulcan* discovered (Baum & Sheehan, 2003; Hanson, 1962).

Or would have been, if it existed.

In the ensuing decades, in Hanson's words, "Vulcan's forge cooled" as various astronomers attempted to make their own observations of Vulcan and failed, or observed spots that appeared promising but later proved illusory. Even had Lescarbault's planet been real, Le Verrier's determinations of its features based on Lescarbault's data indicated that it could not reconcile Mercury's anomalous orbit with NCM. Ultimately, NCM proved unable to explain the anomalous precession of Mercury's perihelion. Instead a radically new theory of gravity – Einstein's General Theory of Relativity – would explain this phenomenon. But that is another story ...

6.4.2 Progress

Lakatos's MSRP is primarily a theory of scientific *progress*: It draws its principal distinction between *progressive* and *degenerating* research programs. Whether a research program is progressive depends on its ability to generate predictions of *novel* facts, i.e. "facts which had been either undreamt of, or have indeed

been contradicted by previous or rival programmes" (Lakatos, 1978, 5). Progress comes in two varieties: A *theoretically progressive* research program is one in which later theories, constructed with the aid of the program's positive heuristic, predict novel facts. In an *empirically progressive* research program such predictions are corroborated by subsequent investigation. A research program in which responses to anomalies are not motivated by the positive heuristic of the research program or do not result in corroborated novel predictions is *degenerating*.

NCM's encounter with the anomalous orbit of Uranus displays its progressive character. By adding a heuristically motivated hypothesis about a new planet, Le Verrier constructed a new theory with novel content, which was then corroborated. The attempt to supplement NCM with a hypothesis about an intra-Mercurial planet was theoretically progressive, but NCM failed to convert this into empirical progress. As Lescarbault's Vulcan proved illusory, astronomers explored other means to solve the problem of Mercury's orbit.

One such effort reveals the challenge of applying Lakatos's MSRP. In 1894 Asaph Hall suggested a revision of Newton's law of gravity itself. He proposed to change the parameter that dictates how the strength of the gravitational force depends on the distance between the two gravitating masses. The constraints on viable revisions were significant, because any such change would have to preserve the very good *agreement* between other astronomical data and the standard formulation of Newton's law (in which gravitational force F_G is proportional to the inverse of the distance r squared, i.e., $F_G \propto r^{-2}$). Hall determined that revising Newton's law such that $F_G \propto r^{-2.00000016}$ would yield the right value for the precession of Mercury's perihelion. As William Harper relates, in 1895 Simon Newcomb, "by then the major doyen of predictive astronomy, cited Hall's proposal as 'provisionally not inadmissible' after rejecting all other accounts he had considered" (Harper, 2007, 937).

Was this a theoretically progressive problemshift, or was NCM at this stage degenerating? That depends on whether modifications to the distance dependence in Newton's law of gravity adhere to NCM's positive heuristic. Research programs do not, however, wear their heuristics on their sleeves. Newton's law of gravitation asserts that $F_G \propto r^{\chi}$, where $\chi = -2$. This expression separates the *form* of the distance dependence from its *parameter values*. Does the parameter value $\chi = -2$ belong to the hard core of NCM, or only the form $F_G \propto r^{\chi}$? This question lacks a clear answer; worse, it is not clear what kind of evidence we would need to answer it.

We do know that Hall's proposal did not survive. In 1903 Ernest William Brown used data on the lunar orbit to show that Hall's proposed change to the distance parameter in Newton's law exceeded by two orders of magnitude the maximum allowable difference from -2. And a departure small enough to be allowed could not account for the anomalous orbit of Mercury (Harper, 2007). Something else had to give, and that something was NCM itself.

6.4.3 Novelty

We have seen that Lakatos's distinction between progressive and degenerating research programs turns on their ability to predict *novel facts*. The novel prediction requirement calls to mind a widely held intuition that it is more impressive when a theory can predict something previously unknown than when it can merely explain what is already known.

As appealing as this intuition might be, when we scrutinize what we mean by calling a fact *novel*, and try to understand *why* such predictions should render theories preferable, things become less clear. For example, Le Verrier's prediction of the existence and location of Neptune strikes us as an impressive success not only for Le Verrier himself, but for the theory of NCM on which he relied. Once the existence of Neptune was established, one could then use knowledge of the parameters of Neptune's orbit to predict where it would be in the night sky on some date in the future. This strikes us as somehow less impressive. Yet in both cases one derives from the theory a statement about some state of affairs that is not known already to obtain.

Both predictions would be novel in the *temporal* sense, according to which a fact is novel if at the time that it is predicted it is not among the facts known to science. But the difference in the import of the two predictions calls for a new distinction. What impresses us in a theory is not just the ability to generate a statement about some state of affairs that is in the future, but rather the prediction of some new *kind* of fact, or phenomenon, previously unknown to science. Le Verrier's prediction of Neptune seems to meet this stronger requirement, whereas the prediction of the location in the night sky of an already known planet on some future date does not.[5]

[5] Nineteenth-century scientist and philosopher William Whewell relied on a roughly similar distinction in his influential account of theory evaluation. He called the ability of a theory to predict new kinds of phenomena the "*consilience* of inductions" and held that

Setting aside the non-trivial problem of determining, in a principled way, what constitute facts 'of the same kind,' this temporal notion of novelty would make the determination of whether a research program is progressing depend rather strongly on what seem to be incidental factors that should not really have a bearing on our assessment of a scientific theory. It also raises thorny questions about what it means for a fact to be 'known to science.'

Moreover, whatever the details of our resolutions of these difficulties, we seem bound to leave the status of a research program hostage to biographical matters regarding scientists whose relevance to the credibility of a theory seems dubious. Would Le Verrier's prediction not count as novel if another astronomer had already observed Neptune? What if numerous astronomers had seen the planet? How many would it take to disqualify the prediction from being novel? If only Le Verrier's state of knowledge mattered, would we have to revise our evaluation of NCM upon the discovery of a diary entry revealing that Le Verrier had secretly already observed Neptune before his calculations? It is far from clear just *why* such biographical details should have a bearing on the status of a theory.

Alan Musgrave discussed these difficulties in a survey of various notions of novelty (1974). Musgrave notes that Lakatos himself articulated a notion of novelty that is not temporal. According to Lakatos's account, as Musgrave interprets and defends it, scientists do not use novel predictions for the confirmation of a theory T in isolation, but evaluate T in comparison to some prior or 'background' theory T_0. For Le Verrier this background theory T_0 would perhaps be constituted by the 'hard core' principles of NCM in conjunction with, among other things, existing knowledge of the first six planets in the solar system. Le Verrier derived his prediction from a new theory (call it 'NCM+Neptune') combining all those elements with a 'theory of Neptune' specifying its hypothetical orbital parameters. Regardless of whether anyone in particular, or even the entire astronomical community, already knew of the existence of such a planet, the theory NCM+Neptune has novel predictive content in comparison to T_0 and it is this fact that makes the episode progressive.

The problems we raised for temporal novelty appealed to intuitions about what kinds of facts ought to be relevant to the scientific status of a theory.

consilient theories such as NCM enjoy a higher grade of support than those that merely predict new facts of a kind they are already known to explain (Whewell, 1847).

Such intuitions could be wrong, or at least misleading. Intuitions often are. In this case, however, some philosophers have found a deeper principle underlying these intuitions. What seems impressive about Le Verrier's prediction of Neptune concerns not simply his prior knowledge or lack of knowledge of the existence of a planet beyond Uranus. What matters is that NCM had conceptual resources to supply this prediction *independently* of any interest in enabling it to do so on the part of those who formulated it. The answer was not *built into*, but rather *found in* the theory.

Such a line of thinking has led to the reformulation of Lakatos's novelty requirement in terms of 'heuristic' or 'use' novelty. In Elie Zahar's account, a fact is novel in the heuristic sense relative to a given hypothesis "if it did not belong to the problem-situation which governed the construction of the hypothesis" (Zahar, 1973, 103). John Worrall has articulated this idea in terms of what he calls "use-novelty": some fact E that is entailed by a theory T cannot serve as a test or "potential falsifier" (in Popper's terms) of T if E was used in the construction of T, i.e., if T was "engineered to entail" E (Worrall, 1989a, 148; see also Worrall, 1985 and, for a recent exchange on the issue, Mayo & Spanos, 2009, 125–169). (Much of the ongoing debate over novelty takes place in a framework no longer tied to Lakatos's methodology of scientific research programs.)

6.5 Conclusion: no "instant rationality"

Suppose that we accept Lakatos's demarcation between progressive and degenerating research programs in terms of the prediction of novel facts (however we wish to interpret that requirement). Having made the demarcation, how should we understand its significance?

Lakatos insisted that scientific methodology is important because it "concerns our central intellectual values." Lakatos would have us value *progress*, i.e., those research programs that change their protective belts (in a manner guided by their positive heuristic), while protecting their hard cores (in keeping with the negative heuristic), and in the process succeed in predicting novel facts which are then observed.

Let us, for the sake of argument, accept this point. What follows from it for our judgment of a particular theory at a particular time? We might be interested in deciding whether there is good reason to think the theory is likely to be correct, or at least reliable for purposes of prediction. A working

scientist might be interested in deciding whether she ought to invest her time in working on a particular theory in the expectation that the theory will be successful in solving problems of interest to her and her colleagues. Can Lakatos's methodology of scientific research programs help with such decisions?

By Lakatos's own insistence, the determination of a research program's status as progressing or not *cannot* by itself answer any such questions. It can only show in retrospect the rationality (or failure of rationality) of such decisions, through what Lakatos calls the *rational reconstruction* of the history of science. A rational reconstruction starts with a certain view of what is rational (for Lakatos, this would be the methodology of scientific research programs, but for a Popperian falsificationist or an inductivist it would be something different), and then uses that theory of rationality to show how scientists were rational in a particular scientific episode, under the tentative presumption that those scientists were rational. Lakatos acknowledges that one must then compare the rational reconstruction of the episode with "actual history" in order to "criticize one's rational reconstruction for lack of historicity, and the actual history for lack of rationality" (Lakatos, 1978, 53).

Such a method will be of no direct help for decisions regarding current, on-going science. Lakatos states explicitly that he does not intend anyone to use his criterion to execute a "quick kill" of a scientific theory, as one might use a Popperian falsification: "One must treat budding programmes leniently: programmes may take decades before they get off the ground and become empirically progressive" (Lakatos, 1978, 6). Neither is it a rule, according to Lakatos, that one should stay with a research program as long as it *is* progressive: "It would be wrong to assume that one must stay with a research programme until it has exhausted all its heuristic power" (Lakatos, 1978, 68). It is starting to seem as though there are no clear, objective rules for scientists to follow.

On the contrary, Lakatos insists, his methodology provides an objective basis for the decision to reject some programs outright: "such an objective reason is provided by a rival research programme which explains the previous success of its rival and supersedes it by a further display of heuristic power," i.e., ability to generate factual novelty (Lakatos, 1978, 69). *Finally*, something clear cut and definitive …

…Except! Lakatos immediately re-complicates things by noting that factual novelty is often not susceptible to quick and unanimous appraisal:

"*the novelty of a factual proposition can frequently be seen only after a long period has elapsed*" (Lakatos, 1978, 69; emphasis in original). This is because, he argues, what appears to be an established fact predicted by an already-existing theory can be regarded as novel if a new theory provides us with a new *interpretation* of that fact. This is why we should be tolerant of "budding" research programs, but even an "old, established and 'tired' programme, near its 'natural saturation point' … may continue to resist for a long time and hold out with ingenious content-increasing innovations even if these are unrewarded with empirical success" (Lakatos, 1978, 72).

Lakatos summarized these nuances thus: "the methodology of scientific research programmes does not offer instant rationality" (Lakatos, 1978, 6). His friend in life and foe in philosophical debate, Paul Feyerabend, took Lakatos's denial of instant rationality to be a way of evading a deeper truth to which Lakatos's insights pointed, but which Lakatos would not accept: there is no universal, exceptionless methodology for distinguishing the rational from the irrational, the progressive from the degenerating, good science from bad science. Feyerabend espoused an "epistemological anarchism" according to which the only universally valid and exceptionless rule of method is "anything goes," and he considered Lakatos to be a "closet anarchist." Not surprisingly, Lakatos rejected this characterization of his own views.[6] It is to Feyerabend that we turn next.

[6] Lakatos and Feyerabend planned to carry out a joint project in which they were to present opposing arguments regarding the rationality of science. Lakatos's death precluded that project, but some idea of what it might have been is conveyed by a fascinating – and sometimes appalling – volume that incorporates some of Lakatos's lectures (fascinating) and the Lakatos–Feyerabend correspondence (this is where you will find the appalling bits, along with displays of great wit; in some instances they are one and the same) (Lakatos & Feyerabend, 1999).

7 Feyerabend: epistemological anarchism

7.1 Introduction

" … a little brainwashing will go a long way in making the history of science duller, simpler, more uniform, more 'objective' and more easily accessible to treatment by strict and unchangeable rules" (Feyerabend, 1988, 11). In the introduction to his book *Against method*, Paul Feyerabend thus expresses his views about the project of the methodology of science pursued by most philosophers (as he understands them). He not only disagrees with their pursuit of this project, he considers it *harmful*.

What is the philosophical project that he rejects? Why does he reject it? What does he offer in its place? In this chapter we will seek answers to these questions. First, however, a word of caution: Feyerabend's willingness to mix what we might think of as reasoned philosophical arguments with polemics makes for entertaining and stimulating reading, but it can also complicate the search for his underlying argument. Moreover, for systematic reasons that we will discuss, Feyerabend sometimes plays 'devil's advocate,' defending a position that is not really his own, and it is not always obvious when he does so.

Like his friend Imre Lakatos, Paul Feyerabend came of age in the midst of the turmoil enveloping central Europe in the 1930s and 1940s.[1] Born in Vienna on January 13, 1924, he was drafted into the German army in 1942, four years after Germany's annexation of Austria. After serving several tours of duty on the Russian front, including commanding a company of

[1] Feyerabend's autobiography is very readable and conveys his humor and breadth of interests (Feyerabend, 1995). Paul Hoyningen-Huene's obituary for Feyerabend includes a biographical sketch and appreciation of Feyerabend's contributions to philosophy of science (Hoyningen-Huene, 2000).

'seasoned soldiers,'[2] he was shot in the spine, leaving him paralyzed from the waist down. He used a wheelchair for a while until he learned to walk with crutches, and throughout his life suffered bouts of intense pain requiring medication.

After the war, Feyerabend returned to Vienna, where his studies included physics and mathematics. After completing a PhD in philosophy in 1951 he traveled to London, where he studied with Karl Popper. His teaching career took him from Bristol to Berkeley to wanderer, lecturing widely and sometimes holding three positions simultaneously on two continents (Berkeley, London, and Berlin). So he had already become well-known and sought after as a lecturer and philosophical interlocutor when he published a long essay titled "Against method: Outline of an anarchistic theory of knowledge" (Feyerabend, 1970). When he expanded that essay into a book with the same title, published in 1975, he went from well-known to notorious.

7.2 Epistemological anarchism

In *Against method*, Feyerabend advocates a view that he calls *epistemological* (or sometimes *theoretical*) *anarchism*. He begins the book by proclaiming, "Science is an essentially anarchic enterprise: theoretical anarchism is more humanitarian and more likely to encourage progress than its law-and-order alternatives" (Feyerabend, 1988, 9).

The term "essentially" deserves emphasis. The anarchic character of science is not, Feyerabend emphasizes, a mere historical contingency. Although Feyerabend will invoke the history of science in his arguments, he is not simply *observing* something about how science has as a matter of fact been pursued, he is *advocating* anarchism as an integral and value-enhancing part of the scientific undertaking.

Next, note that Feyerabend claims *two* distinct advantages of an anarchistic approach to science. First, it is "more humanitarian ... than its law-and-order alternatives." So Feyerabend's epistemological anarchism does not rest only

[2] "There I was, a dedicated bookworm, with no experience, the symbols of authority on my shoulders, confronted with a bunch of skeptical experts. It happened to me again, twenty years later, when I was supposed to teach the Indians, blacks, and Hispanics who had entered the university as part of Lyndon Johnson's educational programs. Who was I to tell these people what to think?" (Feyerabend, 1995, 46–47).

on epistemological considerations in any narrow sense. It is part of a more general view of the conditions for human flourishing, which Feyerabend associates with the nineteenth-century British philosopher John Stuart Mill. Second, epistemological anarchism is "more likely to encourage progress than its law-and-order alternatives." Feyerabend will need to show that his approach is more fruitful in the pursuit of scientific progress than these alternatives.

But what *is* epistemological anarchism? Put simply, it is the position that science should not be pursued by following a fixed set of rules at all times.

This statement is of course purely negative, and the simplicity of its formulation belies the difficulty of determining its consequences. To flesh it out, let us consider carefully Feyerabend's two-fold argument for epistemological anarchism, alluded to above. I will refer to these two main lines of argument as *the argument from progress* and *the argument from happiness*.

7.3 The argument from progress

We can perhaps capture the main point of it with a negative statement: Whatever rules you might think of as constitutive of the scientific method, there will be some circumstances in which one can best advance scientific knowledge not by adhering to those rules, but by violating them. "To those who look at the rich material provided by history," Feyerabend writes, "and who are not intent on impoverishing it in order to please their lower instincts, their craving for intellectual security in the form of clarity, precision, 'objectivity', 'truth', it will become clear that there is only *one* principle that can be defended under *all* circumstances and in *all* stages of human development. It is the principle: *anything goes*" (Feyerabend, 1988, 19; emphases in original).

This expression, "anything goes," has become a kind of icon of Feyerabend's philosophy of science. Plucked out of context, however, this expression can create misunderstanding. In particular, one might think that with this expression Feyerabend tells us that one may advance scientific knowledge by whatever means one likes. This would be a very odd position to take, because it is not hard to see that, to the contrary, most things that one *could* do in science (setting your apparatus on fire, pouring maple syrup into the spectrometer) will not work. Indeed, finding even *one* thing to do that will advance significantly the state of knowledge in a given field requires great ingenuity and effort.

Feyerabend's actual position, though perhaps ambiguous, is certainly more nuanced than just "anything goes." Instead, he argues that any rule that we might regard as a 'universal' guide to scientific inquiry will, under some circumstances, prevent scientists from contributing to the progress of science (whatever we think of as progress).

To support this claim, Feyerabend employs two argumentative strategies. The first employs general philosophical arguments regarding the relationship between rules and their context of application, and regarding the conditions under which arguments function effectively to change opinions. The second strategy employs episodes drawn from the history of science. Feyerabend's writings weave these two strategies together, using history to lend plausibility to his philosophical arguments, and refracting his interpretation of historical events through his own philosophical lenses. To aid us in understanding the arguments, I will discuss the strategies separately. Let us begin with the philosophical arguments.

7.3.1 Limitations of rules and logic

If we knew exactly what the world is like, then perhaps we could come up with a system of rules that would constitute an effective procedure for determining its structure, its constitution, the general principles that best describe how the world works. Of course, in that case we would not really need these rules, because we would already know everything that the rules were designed to help us discover. Feyerabend argues that the effectiveness of a rule for pursuing science will always, in one way or another, depend on what the world is like, which is precisely what we do not know. So we must "keep our options open." "Epistemological prescriptions may look splendid when compared with other epistemological prescriptions, or with general principles – but who can guarantee that they are the best way to discover, not just a few isolated 'facts', but also some deep-lying secrets of nature?" (Feyerabend, 1988, 12).

We have already encountered this idea in previous chapters, though not from this particular perspective. Recall our unscientific doughnut example from Chapter 1: Whether it is reasonable to generalize from the unpleasant taste of a sample of horseradish doughnuts to horseradish doughnuts in general depends in part upon the extent to which the taste attributes of doughnuts are uniform, a fact about which we are imperfectly informed.

Similarly, the reasonability of Newton's generalization about the variable refrangibility of the rays in white light depends on the extent to which light behaves the same independently of location and source.

Nor can one disregard this as an aspect of the problem of induction that we can avoid by using a falsificationist method. Falsificationist rules also depend for their success on contingent and uncertain facts. A falsificationist response to the anomalies in Uranus's orbit would have called for the rejection of Newtonian celestial mechanics on the basis of that theory's failed predictions. That is the right response only if there are no other celestial bodies exerting gravitational forces that render Uranus's orbit consistent with Newton's theory (within the limits of accuracy available at the time).[3]

Such considerations suggest that whether a given application of a rule of inference — either inductive or falsificationist — leads to success will be a contingent matter, depending on additional facts not in general known in advance.

Alright, so science depends on *fallible* methods. Even when we apply the methods correctly (whatever we think correctness might require), the conclusions we draw might be mistaken. That a rule is fallible, however, does not entail that we should not follow it. In particular, if following a methodological rule yields *reliable* results, or results more reliable than those obtained by alternative methods, then it is reasonable for scientists to follow that rule.

Were Feyerabend asserting merely the *fallibility* of widely endorsed methodological rules in this way, we would have little reason to reject his assertion, but also little reason to think of him as an original and provocative philosopher. But Feyerabend goes further. He claims that, for any rule that we might consider (partly) constitutive of the scientific method, there will be occasions when we can *better* advance scientific knowledge by following a rule that points us in the *opposite* direction.

7.3.2 Proliferation of theories

As an example of this, Feyerabend defends the use of *counterinduction*, which "advises us to introduce and elaborate hypotheses which are inconsistent

[3] More precisely, this describes the response of what Lakatos calls a "naïve falsificationist." The sophisticated falsificationist encounters the same problem in a slightly different form.

with well-established theories and/or well-established facts" (Feyerabend, 1988, 20). This bit of advice will strike many as counterintuitive, to say the least. If you consider a certain theory or fact to be well-established, why would you waste your time on a theory that is incompatible with it? That you should not thus waste your time might partly be the point of the last of Isaac Newton's four "Rules for the Study of Natural Philosophy," which states: "In experimental philosophy, propositions gathered from phenomena by induction should be considered either exactly or very nearly true notwithstanding any contrary hypotheses, until yet other phenomena make such propositions either more exact or liable to exceptions." Newton goes on to explain that the point of this rule is "so that arguments based on induction may not be nullified by hypotheses" (i.e., alternative theories introduced for no other reason than their possible ability to explain the phenomena thus far observed) (Newton, [1687]1999, 796). Newton's rule, by making the conclusions of inductive arguments provisional, implicitly acknowledges the *fallibility* of inductive reasoning that we just highlighted, but he does not think that the fallibility of induction makes it reasonable to adopt Feyerabend's counterinductive procedure in its stead.

Newton notwithstanding, Feyerabend believes that introducing incompatible alternatives *is* (sometimes) reasonable, and he gives somewhat different arguments to support this for the case in which the counterinductively introduced theory contradicts a well-established theory and the case in which it contradicts well-established facts.

Incompatibility with well-established theory

First, he argues, even a well-established theory may have weaknesses that will only be discovered if one has some other incompatible theory with which to compare it, since "some of the most important formal properties of a theory are found by contrast, and not by analysis" (Feyerabend, 1988, 21).

In a certain sense, such a statement seems eminently reasonable. There is no doubt that the distinctive features of a given theory can be highlighted by comparison with another theory with which it competes for our acceptance. In the *Origin of species*, Charles Darwin devotes a significant amount of attention to comparisons between the kinds of explanations of traits of species offered by his theory of evolution by means of natural selection and those offered by the theory of "special creation." He thus helps his readers appreciate the differences between the two accounts – a comparison that

remains illuminating long after the scientific evidence for evolution by natural selection has been well-established. That arguments offered in support of one theory often involve a comparison with another (either an already existing competitor, or a foil contrived just for the occasion) is not only not controversial, it is almost too obvious to be interesting.

Reading Feyerabend's elaboration of this point, however, it quickly becomes evident that, whatever his point is, it certainly is *supposed to be* controversial. Consider the following:

> A scientist who wishes to maximize the empirical content of the views he holds and who wants to understand them as clearly as he possibly can must therefore introduce other views; that is, he must adopt a *pluralistic methodology*. . . . Proceeding in this way he will retain the theories of man and cosmos that are found in Genesis, or in the Pimander,[4] he will elaborate them and use them to measure the success of evolution and other 'modern' views. He may then discover that the theory of evolution is not as good as is generally assumed and that it must be supplemented, or entirely replaced, by an improved version of Genesis. Knowledge so conceived is not a series of self-consistent theories that converges toward an ideal view; it is not a gradual approach to the truth. It is rather an ever increasing *ocean of mutually incompatible (and perhaps even incommensurable)* alternatives . . . (Feyerabend, 1988, 21; emphasis in original).

Feyerabend's choice of examples seems outrageous. Is he seriously claiming that we can improve our understanding of biological species by replacing Darwinian ideas with those found in the biblical text of Genesis? As mentioned at the beginning of this chapter, Feyerabend often took on the role of devil's advocate, defending views to which he had no real commitment. This makes it hard to gauge the seriousness of the specific example he gives here. We should not, however, let that distract us from the seriousness of the main point, which concerns the very aim of science itself. Science aims, he tells us, not at a single comprehensive account of the phenomena in a given domain, but at an "ocean" of accounts that are not even compatible with one another.

As a consequence, Feyerabend advises against discarding any idea: "There is no idea, however ancient and absurd, that is not capable of improving our knowledge" (Feyerabend, 1988, 33).

[4] The Pimander, which describes a mysterious cosmological revelation, constitutes one chapter of a collection of second and third century CE Greek texts presenting eclectic and opaque teachings of a spiritual nature called the *Corpus Hermeticum*.

Incompatibility with established facts

Turning to the second argument for counterinduction, which defends the value of introducing theories inconsistent with "experiments, facts, observations," Feyerabend begins by reminding us of an already familiar point: When a theory is introduced, it confronts anomalies from the very beginning: "no single theory agrees with all the known facts in its domain" (Feyerabend, 1988, 39). Feyerabend devotes a chapter of *Against method* to a somewhat breezy catalog of examples of this phenomenon as exemplified in the history of physics, and then turns to a rather detailed discussion of his favored historical illustration of counterinduction: Galileo's arguments in defense of Copernicus's heliocentric theory of the cosmos.

Before turning to Feyerabend's treatment of Galileo, note that Feyerabend's argument *starts* with the rejection of an *unrealistic* methodological constraint. Because of the difficulties that typically face theories at the time of their introduction, "[t]he demand to admit only those theories which are consistent with the available and accepted facts …leaves us without any theory. …Hence, a science as we know it can exist only if we drop this demand" (Feyerabend, 1988, 51). Allowing theories into science that contradict statements that have been accepted as "basic" (to use Popper's terminology) is thus something that we simply have to tolerate if we wish to have new theories.

But Feyerabend's argument does not stop there: we should not merely tolerate such contradiction of the "facts," we should *seek it out*. "Counterinduction is thus both a *fact* – science could not exist without it – and a legitimate and much needed *move* in the game of science" (Feyerabend, 1988, 54; emphasis in original). Such a counterinductive move benefits scientists by allowing them to criticize the facts themselves (i.e., to show that they are not facts after all), in ways that might not otherwise be possible. Feyerabend considers Galileo to have done exactly that.

Galileo on the motion of the earth

In 1632, after a great deal of maneuvering to obtain the needed license from the Church (the *imprimatur*), Galileo Galilei published his *Dialogue concerning the two chief world systems*. The title referred to two competing accounts of the nature and constitution of the cosmos. The *geocentric* (earth-centered) system

was rooted in Aristotle's natural philosophy. Claudius Ptolemy had developed a mathematical model of it in the second century CE, which subsequent astronomers had elaborated into a rather complex system. Nicolaus Copernicus had devised a mathematical model (also rather complicated) of a *heliocentric* (sun-centered) system, which he described in a book titled *De Revolutionibus*, published just as Copernicus died in 1543.

We will not concern ourselves here with the details of the two systems.[5] For our purposes, the significant difference between them is just this: According to the first system the Earth is stationary and the motions that we observe of the planets, the Moon, the Sun, and the stars across the sky are all the result of those bodies following orbits centered on the Earth. According to the second system, the Earth has a two-fold motion consisting of its *diurnal* rotation about its own axis and its *orbital* motion around the Sun.

Galileo's defense of the Copernican system embroiled him in a theological dispute that led ultimately to his trial and to the Church pronouncing him "vehemently suspected of heresy." The Church had put *De Revolutionibus* on the index of prohibited books in 1616, and around the same time Galileo had been personally warned by the Vatican's chief theologian, Cardinal Robert Bellarmine, against teaching or defending Copernicus's system.[6] Galileo used a transparent artifice in the *Dialogue* to claim that it did not defend Copernicanism. Using the dialogue form, Galileo creates a conversation between three figures: Salviati, Sagredo, and Simplicio. Officially, Salviati does not advocate Copernicanism; he intends "only to adduce those arguments and replies, as much on one side as on the other … and then to leave the decision to the judgment of others" (Galilei, 1967, 107). In practice, Salviati presents a relentless stream of arguments directed against the Aristotelian, geocentric viewpoint and in defense of the Copernican system. Galileo uses Sagredo as a neutral arbiter between Salviati and Simplicio, the staunch defender of Aristotelian geocentrism.

Feyerabend discusses many aspects of the *Dialogue*'s extended argumentation. Here I will confine our discussion to one line of argument that illustrates

[5] Thomas Kuhn's first book remains a worthwhile resource for understanding these developments (Kuhn, 1957).

[6] A very large literature has grown up around the Galileo affair. (Finocchiaro, 1989) provides a useful resource, while (Blackwell, 2006) presents an intriguing recent development in scholarship.

especially vividly Feyerabend's strategy of arguing from historical cases for the value of counterinduction.

Imagine, *if you can*, that you live in the pre-Copernican era. Perhaps you have never even heard before of the idea that the Earth might be in motion. You might well find the suggestion of a moving earth positively ludicrous. Indeed, the Copernican system requires the Earth to have two quite rapid motions. We now know that the Earth travels about the sun at approximately 30 km/sec, while the Earth's daily rotation about its own axis entails that locations on the equator travel at about 0.5 km/sec.[7] The magnitudes of these speeds alone suggest that we are in a constant state of considerable motion, and surely, it might seem, such motion would manifest itself to us in *some* noticeable ways.

More specifically, opponents of the Copernican system pointed to a very simple, readily observed phenomenon as conclusive evidence that the Earth *cannot* be moving in the manner required in a heliocentric model. They pointed out that objects that are dropped on earth *fall in a straight line* towards the surface of the earth. The Aristotelian natural philosophy explains this phenomenon naturally: objects such as rocks, bricks, books, and Fabergé eggs fall because they primarily consist of the element earth, the *natural motion* of which is toward the center of the cosmos (which is also why we find the planet Earth situated just there). But, the argument goes, a simple experiment shows such motion to be incompatible with the motion of the Earth about its own axis.

Consider a stone dropped from the top of a tall tower. According to the Copernican theory, the tower is moving very rapidly, along with the rest of the earth, from West to East. Let us suppose the stone is dropped from the western side of the tower. We should then observe the stone to fall away *at*

[7] The former figure would have been estimated much lower in Galileo's day, when estimates of the distance from Earth to Sun were systematically low. In the second century BCE, Aristarchus of Samos had used an ingenious method to arrive at the result that the distance from the center of the Earth to the Sun is about 1528 earth radii (roughly 11,000,000 km). The ninth-century CE Arab astronomer Al Fargani, using a different method, placed the Sun at 1220 Earth radii's distance (about 9,000,000 km) (Kuhn, 1957, 80–81, 274–278). Modern measurements put the figure at about 150,000,000 km. Even with the smaller values, pre-Copernican intuitions would lead one to conclude that the speed at which the Earth would have to orbit the sun must surely be *noticeable*.

an angle from the tower, for once it has been released the tower will continue its motion from West to East, but there is nothing to make the stone keep up with the tower. It will fall behind while it also falls to the ground. But, of course, this is not what we see at all. The dropped stone falls in a straight line, perpendicular to the surface of the Earth.

According to Feyerabend, Galileo's defense of Copernicanism requires him to *replace* the fact 'the dropped stone falls in a straight line, perpendicular to the surface of the Earth' with another fact: 'the *apparent* motion of the stone is described by a straight line, perpendicular to the surface of the Earth' and then to show how this new fact is powerless to refute the Copernican theory, since one can interpret it in terms of either a stationary or a moving Earth. To do this, however, Galileo must get his reader to take a single fact and, in light of a theory with which that fact is incompatible, introduce a new distinction between the sensations experienced and the interpretation given to those sensations. Feyerabend uses the term *natural interpretations* to refer to the mental operations that operate so immediately upon our sensations that we do not easily notice them as producing interpretations of those sensations.

In the present case, the natural interpretation is the move from our sensations of the stone falling along a trajectory parallel to the walls of the tower to the claim that 'the stone's actual motion is described by a straight line, perpendicular to the surface of the Earth.' Galileo's problem is that to his Aristotelian opponents, this is not a move at all. They identify the actual motion of any object with the motion that they observe it to exhibit with the unaided eye. To get his opponents to admit the distinction that renders that observation powerless against the Copernican system requires that Galileo *first* get them to adopt a new perspective on the observation different from and incompatible with Aristotelian epistemology.

In that new perspective, two new principles play a central role. The first of these is the principle of inertia (more accurately, a precursor to the more modern principle of inertia articulated as the first of Newton's laws of motion: a body with a given state of motion will remain in that state of motion if no forces act upon it). Galileo does not state this principle explicitly, but rather has Salviati lead Simplicio through a thought experiment involving a ball rolling on an unbounded frictionless plane, securing Simplicio's admission that such a ball, once set into motion, would have no reason to stop moving. Later, Simplicio is required to apply the same reasoning to the diurnal

motion imparted to the stone by the moving Earth.[8] The second principle is the principle of the relativity of motion. Galileo states this principle as follows: "Motion, in so far as it is and acts as motion, to that extent exists relatively to things that lack it; and among things which all share equally in any motion, it does not act, and is as if it did not exist" (Galilei, 1967, 116).

Neither of these principles has a place in Aristotle's natural philosophy, so how can Galileo get his audience to accept the distinction he requires? Rather than pursue a lengthy excursus, suffice it to say that he relies on thought experiments like that previously mentioned and his own expert manipulation of the dialogue format to have Salviati lead Simplicio (and the rest of his audience) to the following realization: If the Earth does rotate on its axis as the Copernican system requires, then the stone dropped from the tower will not be left behind but will continue to move with the tower subsequent to being dropped, because of inertia. This results in the stone having an actual motion (i.e., the motion that would be observed from a viewpoint external to the earth) that is curved. But, because the diurnal motion is shared by the stone, the tower, and the observer, the principle of relativity entails that the motion will not appear to be curved; only the straight downward motion of the stone toward the Earth will be observed by an observer on the Earth.

A rather obvious understanding of Galileo's argument would be that it shows that what Galileo's opponents thought was a fact was no fact at all, and reminds us to always confine ourselves to arguing on the basis of the sense impressions themselves. But, Feyerabend argues, such an interpretation fails to give adequate consideration to the historical situation in which Galileo made his argument, as well as the role of theories in observation. Natural interpretations are unavoidable, according to Feyerabend. Galileo's opponents were not being unreasonable in taking themselves to have observed the straight line motion of falling stones; they could hardly have understood things differently, given their cognitive situation. The logical empiricist prescription to separate the theoretical and observational vocabularies and purify our descriptions of any theoretical language will not work. The key to scientific advancement is to *multiply interpretations* by introducing new theories that

[8] The fact that we are now dealing with a circular motion means that the principle of inertia does not strictly apply. Galileo either regarded at least some circular motions as satisfying the principle of inertia, or else declined to state explicitly that the principle applies to the stone's 'horizontal' motion as an *approximation* of straight-line motion (Hooper, 1998).

challenge the established facts of any given time. According to Feyerabend, Galileo did just that.

7.4 The argument from human flourishing

The other main line of argument pursued by Feyerabend is that a rule-bound approach to science is not conducive to human flourishing on either the individual or societal level. Educating people about science by advocating strict rules for the pursuit of knowledge "cannot be reconciled with a humanitarian attitude" (Feyerabend, 1988, 12).

Although he argues the point in less detail, Feyerabend regards humanitarian considerations as the primary *motivation* for his epistemological anarchism. He states in the third edition of *Against method* that "Anger at the wanton destruction of cultural achievements from which we all could have learned, at the conceited assurance with which some intellectuals interfere with the lives of people, and contempt for the treacly phrases they use to embellish their misdeeds was and still is the motive force behind my work" (Feyerabend, 1988, 272).

7.4.1 Feyerabend and Mill

Admirable motivations notwithstanding, Feyerabend owes us an argument to warrant the claim that only epistemological anarchism is compatible with a "humanitarian attitude" toward science. We might apprehend that argument by attending to a source that Feyerabend often cited as a precedent for his own views: John Stuart Mill's 1859 essay "On liberty." Mill co-wrote this essay with his wife Harriet Taylor to discuss "the nature and limits of the power which can be legitimately exercised by society over the individual" (Mill, [1859]1963, 213). Mill's thesis and his passionate defense of it have rightly earned the essay both controversy and admiration for its provocative thesis and weighty argumentation.

The thesis deserves to be quoted in full:

> The object of this Essay is to assert one very simple principle, as entitled to govern absolutely the dealings of society with the individual in the way of compulsion and control, whether the means used be physical force in the form of legal penalties, or the moral coercion of public opinion. That principle is, that the sole end for which mankind are warranted, individually

> or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. (Mill, [1859]1963, 223)

The argument in support of this thesis, Mill claims, depends entirely on considerations of *utility*, i.e., of what conduces to the happiness of humanity as a whole. We will confine our attention to aspects of Mill's argument most relevant to Feyerabend's appreciation (or appropriation) of Mill's essay.

Mill devotes one chapter of his essay to a consequence of his thesis that holds great significance for Feyerabend: freedom of thought and expression. Because one's thoughts never, and one's speech only in rare circumstances, pose a threat of real harm to another person, Mill espouses a liberty of thought and expression that is very nearly complete. (He emphasizes that to offend a person is *not* to harm him.) Moreover, Mill aims to show (in keeping with his emphasis on utilitarian considerations) the great *benefit* to society of the expression of opinions that run counter to popular belief or to what is encouraged by those wielding social or cultural power and influence.

He summarizes his view as follows: "the peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right they are deprived of the opportunity of exchanging error for truth; if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error" (Mill, [1859]1963, 229).

Mill considers separately the cases in which an opinion that is subject to suppression is true and in which it is false. That true opinions ought not be suppressed seems plausible enough, but one might think that we are well justified in using the power of "moral coercion" against views that are clearly incorrect. But, Mill notes, given human fallibility, we cannot be certain that what we regard as the truth really *is* the truth. Suppressing what we *believe* to be false might inadvertently amount to suppressing the truth.

But Mill goes further: even when an opinion *really is* false, we benefit from its expression, for two reasons. First, we gain a better appreciation for the *reasons* that support a true opinion when we have to defend it against arguments in favor of a contrary view. Second, we gain a better understanding of the *meaning* of an assertion when we encounter arguments to the contrary that we then must defeat.

Feyerabend appreciates that Mill's argument is based on the requirements of increased happiness. He writes that Mill advocates the proliferation of contrary points of view "as the solution to a problem of *life*: how can we achieve full consciousness; how can we learn what we are capable of doing; how can we increase our freedom so that we are able to decide, rather than adopt by habit, the manner in which we want to use our talents?" (Feyerabend, 1981, 67).

Insofar as Mill advocates the expression of views contrary to those held by the many or by the powerful, on the grounds that active engagement with such views benefits all, Feyerabend's citation of Mill seems apt. But has Feyerabend overreached in assimilating Mill's views to his own? Laura Snyder has pointed out that Mill, unlike Feyerabend, does not embrace the proliferation of views as an end in itself, but rather as a *means* to the discovery of truths that might otherwise not be appreciated (Snyder, 2006, 214). He also regards it as a means to the appreciation of the reasons supporting those truths and even the very meaning of those truths.[9]

This last point, however, is crucial for understanding an important aspect of Feyerabend's appreciation of Mill. Mill goes so far as to claim that, when an opinion has fallen so far out of favor that it no longer has supporters, a 'devil's advocate' ought to be assigned the task of defending it nonetheless, so that the rest of us may benefit from having it brought to our attention. Feyerabend himself seems often to assume this very role, by his own admission. As Elisabeth Lloyd has pointed out, reading Feyerabend in the role of Mill's devil's advocate can help us to understand some of his more outrageous claims (Lloyd, 1997).

Recall how Feyerabend suggests that "the theory of evolution is not as good as is generally assumed and that it must be supplemented, or entirely replaced, by an improved version of Genesis." On the face of it, this appears scientifically irresponsible. Feyerabend offers no argument indicating a

[9] Snyder, disputing Feyerabend, as well as my own (1999b) and Elisabeth Lloyd's (1997), rightly points out the difference between Feyerabend and Mill with regard to the status of proliferation of opinion as an end, citing passages in which Mill states that "cessation … of serious controversy" is a necessary concomitant of human progress. However, she understates the importance of the proliferation of *arguments* by failing to note the immediately following sentences in which Mill states that, although such consolidation of opinion is "at once inevitable and indispensable, we are not therefore obliged to conclude that all its consequences must be beneficial," and proceeds to argue for the value of the devil's advocate role (Mill, [1859]1963, 250–251).

defect in the 'modern synthesis' (of the Darwinian idea of evolution by means of natural selection and genetic theory) which could be remedied by this "improved Genesis." The bare possibility that there *might* be some such flaw is insufficient to make his point. Elsewhere, Feyerabend defended astrology, when the *Humanist* published a statement denouncing it, signed by 186 "leading scientists," including 18 Nobel laureates. Feyerabend gives a few reasons to think that it is plausible that "celestial events such as the positions of the planets, of the moon, of the sun influence human affairs" and accuses the denouncers of being ignorant of the history of astrology and even of the relevant science (Feyerabend, 1978, 91–96).

Can Feyerabend really mean to advocate for astrology? Well, yes and no. In a later work he acknowledges that "Astrology bores me to tears," but goes on to note that "it was attacked by scientists, Nobel Prize winners among them, without arguments, simply by a show of authority and in this respect deserved a defense" (Feyerabend, 1991, 165). Here Feyerabend makes it clear that, on the one hand, he really does consider the attack on astrology to be ill-informed, but on the other hand, he does not really think much of astrology. To advocate on behalf of views for which one has no deep conviction is the mark of a 'devil's advocate.'

7.5 Conclusion

If the main point of Feyerabend's philosophy of science were simply that for any proposed rule of method, it is always *possible* that one might better advance science by breaking it than adhering to it, then it is not clear why anyone should get very excited about it. In particular, this point is not strong enough to support any stronger claim about the benefits of proliferating theories than that by doing so it is *possible* that one might come up with a better theory than one already has. The fact that a certain strategy *might* work is not a good enough reason to pursue that strategy. If my car won't start, I *might* fix it by randomly loosening and tightening bolts and clamps, but that does not make this an advisable strategy for fixing cars. Peter Achinstein has argued that if we try to interpret Feyerabend's proliferation thesis in a stronger form it becomes "absurd." Specifically, Achinstein proposes the following way of giving Feyerabend's position "some teeth":

> Believe or accept an "invented and elaborated" theory as true, or probable, or as a good predictor, or as good in some way, if it is inconsistent with the

> accepted point of view, even if the latter should happen to be highly confirmed. (Achinstein, 2000, 38).

As Achinstein notes, this proposal cannot be taken seriously, first, because it asks us to believe or accept *too many* theories (there is no limit to the number of theories one can cook up if one does not care whether they have any evidence or plausibility) and, second, because the mere fact of being inconsistent with an accepted theory or point of view does not by itself constitute any *reason* to believe or accept a theory.

I think we might better appreciate Feyerabend's point – and its limitations – by revisiting his frequent citations of Mill. Mill, certainly, did not advocate believing or accepting theories for the reason that they disagree with accepted points of view. He advocated *listening* to the arguments of those advocating claims that conflict with what is accepted, and even, if no one is willing to take up the cause of an abandoned point of view, having those who do not believe or accept such claims present arguments on their behalf anyway. Similarly, I read Feyerabend as recommending simply that scientists (and others) be willing to try out *some* ideas that conflict with accepted theories (even well-supported ones), so that one can find out how good those ideas can be made. However, this bit of methodological advice is at best suggestive, leaving the question of *which* ideas to pursue completely open. Thus, Feyerabend's philosophy of science seems to have a large gap in its *positive* recommendations.

This gap brings us back to Mill. An important aspect to Mill's philosophy that Feyerabend neglected was his positive contributions to scientific methodology. Although "On liberty" figures among Mill's most influential writings, he devoted much more of his life to a weighty tome titled *A system of logic*, which went through eight editions between 1843 and 1872. Prominent in this work is Mill's articulation of a set of methods for discovering causal relationships. ("Mill's methods" are widely taught to this day.) Although Mill's advocacy of methods of discovery might seem incompatible with the "against method" position Feyerabend finds in "On liberty," I have argued elsewhere that a careful reading of *A system of logic* reveals Mill's nuanced understanding of the limitations on the rules of method that he advocates (Staley, 1999b). In short, Mill saw no conflict between recognizing that *no rule of method is universal and exceptionless* (just as Feyerabend insists) and advocating that nonetheless *some rules of method can be helpful and usually trustworthy*.

Ultimately, the value of Feyerabend's contributions to philosophy of science may lie as much in the way he made them as in their content. His provocateur's stance was not simply a reflection of his personality, it also reflected his approach to philosophy as a whole. He wrote not to express his convictions, but to provoke others to question theirs. Consider the following passage from a dialogue that he published in 1991:

> A: Are you an anarchist?
> B: I don't know – I haven't considered the matter.
> A: But you have written a book on anarchism!
> B: And?
> A: Don't you want to be taken seriously?
> B: What has that got to do with it?
> A: I do not understand you.
> B: When a good play is performed the audience takes the action and the speeches of the actors very seriously; they identify now with the one, now with the other character and they do so even though they know that the actor playing the puritan is a rake in his private life and the bomb-throwing anarchist a frightened mouse.
> A: But they take the writer seriously!
> B: No, they don't! When the play gets hold of them they feel constrained to consider problems they never thought about no matter what additional information they may obtain when the play is over. And this additional information is not really relevant …
> A: But assume the writer produced a clever hoax …
> B: What do you mean – hoax? He wrote a play didn't he? The play had some effect, didn't it? It made people think, didn't it? (Feyerabend, 1991, 50–51)

Institutionalized science carries a great deal of cultural authority, and Feyerabend, while valuing scientific inquiry a great deal, sought to make people think carefully about the basis for that authority, and even more carefully about the devaluation of beliefs denied the imprimatur of scientific institutions. Surely, such a position can be taken too far, if it causes us to fail to realize the potential benefits of scientific knowledge. But Feyerabend insisted that beliefs could be good in ways not recognized by the standards adopted by scientific institutions, that such beliefs might hold great value for those who hold them, and that it is unwise to fail to understand why simply because one has not taken the time to give those beliefs a fair and knowledgeable hearing.

Part II

Ongoing Investigations

8 Reasoning with probability: Bayesianism

8.1 Introduction

The last few chapters have presented a series of challenges to the idea that disputes over the status of scientific theories can be resolved unambiguously by appealing to the evidence. We have encountered arguments that purport to tell us that evidence in some way or another *underdetermines* the choice of theories, that the very meanings of the words used to describe the evidence may incommensurably depend on the theory one chooses when interpreting the evidence, that different scientists will disagree about the value of various features of the evidence, that even when the evidence seems to point strongly against a theory, one may rationally hang on to that theory, and even that for any given body of evidence, opposing conclusions based on incompatible methodological rules can equally advance scientific knowledge.

Taken together, such arguments appear to throw a rather large quantity of cold water on whatever embers of hope may remain for the logical empiricists' project of providing a neutral framework that scientists could use to evaluate the degree to which any given body of data supports or confirms any given hypothesis.

Many contemporary philosophers of science, however, have sought an alternate route by which they might pursue something like this project. Moreover, they have sought to do this in a way that might accommodate some of the post-positivist insights discussed in earlier chapters. The key to the approach they have adopted lies in a certain way of thinking about *probability*.

Probability theory is a well-developed area of mathematics, but statements that use the language of probability lend themselves to different interpretations. In this chapter and the next we will consider two approaches to scientific reasoning that invoke two quite distinct interpretations of

probability. The approach we consider here, the *Bayesian approach* (the name derives from the Rev. Thomas Bayes, whose posthumously published essay (1763) included the eponymous theorem that lies at the heart of the Bayesian approach) attempts to build a theory of *rational belief* on the mathematical framework of probability theory. Although we will emphasize the conceptual underpinnings and consequences of this approach, a satisfactory appreciation for the theory (as well as for the rival error-statistical approach discussed in the next chapter) cannot be achieved without a little understanding of some basics of probability theory. So let us start with a brief tour of the probability calculus. A little very simple algebra and elementary set theory is all that is prerequisite to the mathematics that follow.[1]

8.2 Probability theory

The mathematics of probability theory was developed in the seventeenth century, when it was used to solve problems regarding games of chance. Here is a very simple game involving a standard six-sided die:

Game 1: A single die is tossed one time. If it lands with the 4 facing up, you win. Otherwise, you lose.

It would be helpful to you, especially if you plan on wagering on the outcome of this game, to know the probability that you will win. This probability question seems easy to answer: If the die is fair, then there is a one-in-six chance of it landing with the 4 facing up, and hence of you winning. One hardly needs a mathematical theory to figure that out.

More complicated games require more sophisticated treatments. For example, suppose you are given a choice between two games.

Game 2: A single die is tossed four times, and you win if it lands with 6 facing up at least once.
Game 3: A pair of dice is rolled 24 times, and you win if the dice come up with both dice showing 6 at least once.

[1] The requirements of a brief and mathematically shallow treatment entail that we will have to gloss over some important conceptual issues that would emerge from a more rigorous discussion (see Cox & Hinkley, 1974; De Finetti, 1972; Kolmogorov, 1950; Savage, 1972). For a good philosophical introduction to probability and induction, see (Hacking, 2001), which influenced my own presentation. Howson and Urbach (2005) introduce the topic in the guise of a vigorous defense of personalist Bayesianism.

Which of these games should you prefer to play, if you wish to maximize your chance of winning?

The latter problem is known as 'Chevalier de Méré's problem,' named for the French nobleman with a penchant for gambling his unearned wealth who, according to legend,[2] posed the problem to Blaise Pascal, a pioneer of the mathematics of probability. De Méré had an intuition (correct, as it turns out) that the first game would give him a higher probability of winning than the second. But actually demonstrating that this is the case requires having a set of rules for determining probabilities.

8.2.1 Models: representing the problem

First we need a way of representing the problem we want to solve. For that purpose mathematicians rely on the concept of a *random variable*. For our purposes, an intuitive understanding of this concept will suffice: A random variable is a variable capable of taking on various values (either actual or potential), such that a probability distribution can be assigned to those values. A random variable might be such that it can take on only *discrete* values (like the six possible outcomes for the number showing on a single die), or it might be *continuous* (like a random variable for the heights of students in a given classroom). For Game 1 we might choose to represent the outcomes of the first game using a single discrete random variable X_1 (following standard practice, we will use upper case letters for random variables and lower case letters for the observed values of those random variables). The possible values of X_1 (the *outcome space* Ω) are given by the set $\{1, 2, 3, 4, 5, 6\}$. The outcome space for any process (game, experiment, etc.) is a set consisting of *all possible distinct outcomes* for that process. The question of how probable it is that you will win Game 1 then amounts to asking for the value of p in the equation $Pr_M(X_1 = 4) = p$, which we can read as 'the probability that $X_1 = 4$ is equal to p.' (What is the subscript 'M' doing there? Good question! Hang on for a few more sentences and we will find out.)

The mathematical theory of probability thus gives us a means of representing the question we want to ask, and you might think it would be nice for it also to provide us with a means of answering that question. However,

[2] Legend does not even attribute the correct problem to the Chevalier. Ian Hacking provides a corrective to legend (Hacking, 1975, ch. 7).

the theory of probability is of no help in answering this question. It does not give us a set of principles from which we can deduce the answer. Fortunately, we do not need the theory of probability for the answer. If the die is really fair, the answer is $p = 1/6$, and we are just going to assume that the die *is* fair. More precisely we will suppose that

$$p_i = 1/6 \qquad \text{for } i = 1, 2, 3, 4, 5, 6 \tag{8.1}$$

where $p_i \equiv Pr_M(X_1 = i)$. The probability is thus *uniformly distributed* across the outcome space.

It will at this point seem that the theory of probability has done precious little for us beyond introducing some distracting notation, since you did not need any of what I said in these last few paragraphs to know that the probability of getting a '4' on a single throw of a fair, six-sided die is 1/6. The practical usefulness of the theory will become more apparent when we tackle less obvious questions such as that posed by the Chevalier.

Philosophically, however, we have already taken a significant, though easily overlooked step, which has to do with that subscript '*M*'. In our attempt to answer the question of the probability of Game 1, we have had to make some assumptions. Some of these have already been alluded to explicitly, such as the assumption that the die is fair, and that the die is tossed just once. Others may be less obvious, such as the assumption that no player is allowed to veto the outcome of a toss. Others may seem too obvious to state explicitly, being essential to the very possibility of playing Game 1, such as the assumption that the game is played where there is sufficient gravity to make the die land with a determinate side facing up (or some other mechanism to ensure definite outcomes). As diverse as these assumptions may be, they all show up in one place in our probability calculations: A *model* of the problem we are trying to solve. We derive the answers we give to questions about probability from mathematical models that encode all of the assumptions about the scenario under consideration that are relevant to the probabilities that interest us. For this reason, particular probability assignments here will often be given with a subscript denoting the name of the model from which that assignment is derived. The model *M* that we just used can be thought of as consisting of the random variable outcome space $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the uniform distribution assumption $Pr_M(X_1 = i) = Pr_M(X_1 = j)$ for all $i, j \in \Omega$. (The philosophical point having been made, we will sometimes omit the subscript when the model being used is obvious.)

We may now distinguish two importantly different questions: (1) What is the value of the probability that I am trying to determine, *as given by the model I am using*? (2) Is the model that I am using *adequate* for determining the probability that interests me? One may arrive at a perfectly correct answer to a question of type (1), but find that the answer is useless because the answer to the corresponding type (2) question is 'no.'

8.2.2 Axioms of probability theory

For purposes of calculating odds in games of chance, it makes sense to think of probabilities as mathematical functions whose domains consist of sets of *events*, which in turn we can represent in our model in terms of the values taken on by random variables. In Game 1 an event might be 'the die lands with 3 facing up' and we would represent that with $X_1 = 3$. (We will shortly extend our conception so that we can think of probabilities as applying to sentences.) We will also count as events outcomes with more complicated descriptions, such as 'the die lands with either 3 or 5 facing up.' We want probabilities to apply to these as well.

So it will be helpful to think of the domain on which the probability function is defined as constructed from set-theoretic combinations of the elements in the outcome space, such as $\{X_1 = 3\} \cup \{X_1 = 5\}$, i.e., the *union* of the sets $\{X_1 = 3\}$ and $\{X_1 = 5\}$. Let's call the set S of *all* such possible combinations, for a given problem, the *outcomes set* for that problem.

Now we can characterize a probability function as a function Pr on the set of all subsets of the outcomes set S, with a range consisting of the closed interval $[0, 1]$, such that the following *axioms* are satisfied:

Axiom 1 $Pr(S) = 1$

Axiom 2 $Pr(A) \geq 0$, *for any event* $A \subseteq S$[3]

Axiom 3 *Suppose that* $A_1, A_2, \ldots, A_n, \ldots$ *constitute a (possibly infinite but) countable sequence of events that are mutually exclusive (i.e., for any* $i, j = 1, 2, \ldots, n, \ldots$, *such that* $i \neq j$, $A_i \cap A_j = \emptyset$*). Then:*

$$Pr(A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots) = Pr(A_1) + Pr(A_2) + \cdots + Pr(A_n) + \cdots \tag{8.2}$$

[3] Strictly, the domain of Pr should be defined in terms of a *field* constructed from events in S. See, e.g., (Spanos, 1999, 45–69) for details.

Expressed more intuitively, Axiom 1 states that a probability function assigns the number 1 to the event consisting of 'any of the possible outcomes occurs'. Axiom 2 states that a probability function assigns a number greater than or equal to zero to every possible outcome. Axiom 3, often called the axiom of *countable additivity*, tells us how to determine the probability of disjunctions of possible outcomes for the case where such outcomes themselves are mutually exclusive (there is no possibility that they can co-exist), given the probabilities of the disjuncts: The probability of the disjunction is equal to the sum of the probabilities of the disjuncts considered individually. So, the probability (based on the model M) of the event $\{X_1 = 3\} \cup \{X_1 = 5\}$ (the die lands with either 3 or 5 facing up) is equal to the sum $Pr_M(X_1 = 3) + Pr_M(X_1 = 5)$.

These axioms form the foundation for probability theory (Kolmogorov, 1950) and suffice to define what a probability function is. However, they do not tell us how to handle probabilities in cases where the probability of an outcome of some process depends on the outcome of some previous iteration of that process. (For example, we could use it for calculating the probability of drawing a certain card or combination of cards from a deck if after each draw we replace the card and reshuffle; we cannot yet use it to determine the probability of drawing three consecutive spades without replacing cards.) It is useful, then, to add to our axioms a further definition (we may also consider it a fourth axiom):

Definition 2 (Conditional probability)

$$Pr(A|B) \equiv \frac{Pr(A \cap B)}{Pr(B)},$$

where $A, B \subseteq S$ and $Pr(B) \neq 0$.

This rule does two things for us. If A is *probabilistically independent* of B (i.e., $Pr(A|B) = Pr(A)$), then we can derive a rule for calculating the probability that A and B both happen that does not directly involve conditional probabilities: $Pr(A \cap B) = Pr(A) \times Pr(B)$ (the *rule of multiplication*). But it also tells us how in general to handle cases in which outcomes are probabilistically dependent upon one another, which will be important for understanding Bayesian reasoning.

8.3 Probability ideas and science

Of course, we are looking for more than just a set of rules for working with probabilities. We want to see how ideas about probability might help us make

sense of the role that evidence plays in scientific inquiry. In this chapter and the next we will pursue two avenues of thought that aim to provide a reasoned link between the data that scientists produce and the conclusions that they draw. These two avenues exploit quite different ideas about what probabilities represent.

In the opening passages of his book *The emergence of probability*, Ian Hacking writes,

> Probability has two aspects. It is connected with the degree of belief warranted by evidence, and it is connected with the tendency, displayed by some chance devices, to produce stable relative frequencies. (Hacking, 1975, 1)

In this chapter, we consider the *Bayesian* approach to scientific reasoning, which exploits the 'degree of belief' aspect of probability. The *error-statistical approach*, discussed in the next chapter, exploits the 'relative frequencies' idea.

Thinking of probabilities in terms of relative frequencies might seem to be the most natural way to think about the kinds of examples considered so far. If pressed to explain what it means to say that the probability of winning Game 1 is 1/6, we might give the following rough account: If you play Game 1 over and over again, then, on average, you are going to win about one out of every six games (we will refine this rough parsing in the next chapter). But, we can take that idea and use it to introduce the 'degrees of belief' idea. Once introduced, we will find that the 'degrees of belief' idea can be applied to many situations where relative frequencies are simply inapplicable.

8.4 Bayesian probability

8.4.1 Prizes

Let us return to the choice between Games 2 and 3. Knowing the rules for calculating probabilities, one can establish that the probability of winning Game 2 is $671/1296 \approx 0.518$, while the probability of winning Game 3 is ≈ 0.491. (Can you figure out why? The solution is explained in the appendix to this chapter.) Supposing that the same desirable prize is given for winning either game (a box of delicious horseradish doughnuts, perhaps), that there is no cost (or no difference in cost) to playing either game, but that you may

play only one, it is clear that you should *prefer* playing Game 2 to Game 3. So knowing the probabilities of such outcomes, in combination with some information about the relative values you assign to those outcomes, can lead to conclusions about what preferences you should have.

This suggests a procedure by which, reversing things, we start with your preferences and arrive at conclusions about how probable you *believe* the various possible outcomes of a process are. For example, I might offer you a choice between receiving a valuable prize if you win Game 1 and receiving the same prize if it rains tomorrow. If you choose to receive the prize on the condition that it rains instead of on the condition of winning Game 1, this indicates that you consider rain to be *more probable* than winning Game 1, i.e., $Pr_{you}(\text{Rain}) > Pr_{you}(\text{winning Game 1})$.[4] We can further constrain the range of values for your probability assignment by offering more choices. If you are offered the additional option of receiving the same prize for winning Game 2, and prefer that option to the option of receiving the prize if it rains, then we can conclude that $Pr_{you}(\text{Rain}) < Pr_{you}(\text{winning Game 2})$. And if you know the values of the probabilities of winning in those two games, we can conclude that $1/6 < Pr_{you}(\text{Rain}) < 671/1296$.

Using that procedure requires us to attribute to you some beliefs about probabilities of other outcomes that we can use as a reference. By shifting from prizes to *wagers*, we can eliminate such assumptions, and go directly from your beliefs about the odds at which a bet on a proposition is fair to the probability you must attribute to that proposition in order to satisfy what Bayesians consider a requirement of rational belief. First, a bit more terminology.

[4] The subscript in '$Pr_{you}(P)$' does not mean exactly the same thing as it does in '$Pr_M(A)$'. In the latter, it indicated that the probability value in question was that assigned to the event A by the model M. In the former, the subscript denotes an agent ('You') capable of forming and maintaining beliefs (an 'epistemic agent' in philosophical jargon), and the probability value is that which, in virtue of that agent's beliefs, she attributes to the proposition P. We might relate these two cases in the following way: In cases involving personal probabilities, the subscripts on the probability functions refer to models of the beliefs of the epistemic agents in question, instead of models of chance mechanisms. Notice now that we should therefore regard the symbol 'Rain' in '$Pr_{you}(\text{Rain})$' as denoting the *proposition* 'It will rain tomorrow' rather than the *event* of rain occurring tomorrow.

8.4.2 Wagering

We may think of a wager as an agreement between two parties, S and T, regarding a proposition P. If S 'bets on P' and T 'bets against P', that means that S pays an amount of money X (S's *bet*) for the opportunity to win a sum $X + Y$ (for a net gain of Y), subject to the condition that S wins that amount if and only if P is true. T pays the amount Y (T's *bet*) for the opportunity to win the sum $X+Y$ (a net gain of X) subject to the condition that T wins that amount if and only if P is false. The quantity $X + Y$ is known as the *stake*. S's *betting rate* is equal to the ratio of her bet X to the stake $X + Y$.

Example: Samiksha and Tatyana

Suppose that Samiksha and Tatyana agree on a bet such that Samiksha will pay Tatyana \$5 if Tatyana can eat 10 saltine crackers in 60 seconds without drinking any liquid. If Tatyana cannot do this, then Tatyana will pay Samiksha \$5. Let *Salt* stand for the proposition that Tatyana succeeds at this task. Then Tatyana's betting rate on *Salt* of $\frac{\$5}{\$10} = 0.5$ is equal to Samiksha's betting rate against *Salt*.

We might be tempted at this point to conclude from this that they both consider Tatyana's success to have a probability of one-half. But people who engage in gambling (especially those in the gambling business, like casinos) strongly prefer to take on wagers that give them an *advantage*. For example, Samiksha may have seen lots of other people attempt the saltine cracker trick and fail, and she may decide on this basis that the probability of Tatyana succeeding is in fact much less than one-half, i.e, $Pr_{Samiksha}(Salt) << 0.5$.

Notice that if that were the case, it would not be reasonable for her to switch to the opposite side of the bet at the same betting rate of 0.5, because that would, according to her own assessment, put her at a disadvantage. If she considers the probability of Tatyana's success to be very low, she should only accept a bet *on* Tatyana's success if the payoff is large enough relative to her own bet to make the wager at least fair (and preferably advantageous) for her.

Whereas real-life wagering typically involves the pursuit of opportunities to bet at *unfair* rates, a prominent Bayesian approach relates probabilities to beliefs by construing personal probabilities in terms of *fair betting rates*. In our example, suppose the following holds: If offered the opportunity to bet *on Salt* at a rate of 0.5, Samiksha would prefer to bet *against Salt* rather than to bet *on*

it. If the betting rate *on* is 0.1, she would prefer to bet *on*. But suppose we find that at a betting rate *on* of 0.3, she is indifferent as to which side of the bet she would take.[5] We then call this Samiksha's *fair* betting rate: she does not consider such a bet to offer either side an advantage.

For *personalist* (also called *subjective*) Bayesians, a person's degrees of belief in a set of propositions should be understood as the betting rates that they consider fair for those propositions, i.e., as the rates at which, were one (hypothetically) to agree to wagers regarding them, one would be indifferent between wagering on or against them. Moreover, the personalist Bayesian holds, a person's degrees of belief are collectively *rational* just in case they are probabilities, which is to say that they satisfy the axioms of the probability calculus given above. Sets of beliefs that meet this requirement are by definition *coherent*.

8.4.3 Rationality

The Bayesian approach encompasses both a view about probability and a view about rationality, but Bayesians differ about how to connect these two ideas. They agree on this much: To have beliefs that fail to satisfy the probability axioms (that fail to be coherent) is to be (in some way) irrational. The question is, what makes this so? Here we will content ourselves with a quick, somewhat rough, description of two main positions, which we might label the *pragmatic* view and the *logical* view.

According to the pragmatic view, what is irrational about holding beliefs that fail to satisfy the probability axioms is that one is then subject to what is sometimes called a *Dutch Book*,[6] or a *sure-loss contract* (De Finetti, 1964[1937]; Hacking, 2001, 169).

To refer back to our example, suppose that Tatyana asks Samiksha what she considers to be a fair betting rate on *Salt*, and Samiksha replies that she

[5] Note the use of the subjunctive "would" in the three preceding sentences. We do not wish to assume in general that fair betting rates only exist for agents engaged in actual wagering. A person may, for example, have views about what would constitute fair betting rates, but refuse to enter into wagers for reasons of moral or religious principle.

[6] The origins of this term remain a mystery. Hacking, in his generally excellent (2001, 169), claims to find the term used "in passing" in (Ramsey, 1950), but this is incorrect, as pointed out by Humphreys in his (2008), which discusses the history of the term and its obscure origins.

considers a rate of 0.3 to be fair. Asked what she considers a fair rate on ¬*Salt*, Samiksha says that 0.8 would be fair. Samiksha's fair betting rates for these two propositions clearly cannot be construed as probabilities. *Salt* and ¬*Salt* constitute exclusive alternatives, so any probabilities assigned to them must satisfy **Axiom 3**: $Pr(Salt \vee \neg Salt) = Pr(Salt) + Pr(\neg Salt)$. But by **Axiom 1**, $Pr(Salt \vee \neg Salt) \leq 1$. These two requirements cannot jointly be satisfied if $Pr(Salt) = 0.3$ and $Pr(\neg Salt) = 0.8$.

This suffices to show that Samiksha's fair betting rates in this case cannot be considered probabilities (they violate the rules that probabilities must obey), but does not yet tell us what is irrational about holding such beliefs. According to the pragmatic view, the problem lies in the fact that, according to her own beliefs, Samiksha should be willing to take either side of both a wager on *Salt* and a wager on ¬*Salt*, provided that the wagers take place at the rates that she considers fair. So she should be willing to let Tatyana decide which side she will take. This allows Tatyana to arrange, with Samiksha's compliance, two wagers – one on *Salt*, the other on ¬*Salt* – such that no matter what happens, Samiksha will suffer a net loss, and Tatyana will enjoy a net gain.

For simplicity, we assume that in both wagers, the stake is \$10. In the wager concerning *Salt*, Tatyana instructs Samiksha to bet *on Salt*, while Tatyana bets *against*. The rate for a bet *on* is 0.3, so Samiksha puts up \$3 against Tatyana's \$7. In the wager regarding ¬*Salt*, Tatyana has Samiksha bet *on* ¬*Salt*. The rate for a bet *on* is 0.8, so Samiksha puts up \$8 while Tatyana bets \$2. That Samiksha will lose money whether *Salt* turns out to be true or false can be seen in Table 8.1.

Clearly Samiksha is destined to lose money if she agrees to this combination of bets, although she regards them as fair. Does that make

Table 8.1 *A sure-loss contract for Samiksha (S). Tatyana (T) wins no matter what.*

	subject of wager					
	Salt (S bets on)		¬*Salt* (S bets on)			
	S's gain(loss)	T's gain(loss)	S's gain(loss)	T's gain(loss)	S's net	T's net
Salt true	\$7	(\$7)	(\$8)	\$8	(\$1)	\$1
Salt false	(\$3)	\$3	\$2	(\$2)	(\$1)	\$1

her irrational? And even if it does, how are such examples having to do with wagering preferences relevant for beliefs about scientific questions, or for that matter any beliefs on which one is not contemplating a wager?

Let us take the second question first. The pragmatic Bayesian emphasizes that analyses framed in terms of wagers are simply idealizations of real-life decisions generally. Frank P. Ramsey, whose essay "Truth and probability" is among the foundational works in the pragmatic Bayesian tradition (though it holds hints of the logical approach as well), states this clearly. Ramsey comments that his account of degrees of belief "is based fundamentally on betting, but this will not seem unreasonable when it is seen that all our lives we are in a sense betting. Whenever we go to the station we are betting that a train will really run, and if we had not a sufficient degree of belief in this we should decline the bet and stay at home" (Ramsey, 1950, 183). For Ramsey, measuring degrees of belief concerns beliefs "*qua* bases of action" (Ramsey, 1950, 172), and in choosing to act, we are concerned with the question of gains and losses, not of course in just the monetary sense, but in the broad sense in which we value the possible outcomes of our choices differentially. Of course, scientific beliefs may also serve, in principle, as bases of action.

Somewhat more vexing is the question of why being subject to a sure-loss contract should by itself constitute a breach of rationality. Although few people prefer to lose money, doing so is not in itself irrational. Even if we turn away from strictly monetary wagers, and think of sure-loss contracts as stand-ins for a broader class of doomed undertakings, embarking on something that is guaranteed to fail to produce results that you desire is not *obviously* a failure of rationality with regard to one's beliefs. Moreover, as Alan Hájek has pointed out, for every sure-loss contract that a person is subject to on account of non-coherent beliefs, there is also a sure-*gain* contract (Hájek, 2008). Were Tatyana to feel charitable toward Samiksha, she could insist that Samiksha bet *against* both *Salt* and ¬*Salt* at the above rates, and thus guarantee that Samiksha would enjoy a net profit, regardless.

Even accepting that it is irrational to agree to a sure-loss contract, it does not *immediately* follow that it is irrational to hold beliefs that violate the probability calculus, as Henry Kyburg argues (Kyburg, 1978). Recall that Samiksha's incoherent betting odds only lead her into a Dutch Book when she agrees to combine bets in a certain way. Surely she can avoid this undesirable consequence by declining to accept a bet of this kind. The connection between avoiding Dutch Books and rationality might be re-established by insisting

that a rational person must be willing to accept combinations of bets, but it is hard to see on what basis such a principle might be defended.

The *logical approach* to connecting probability and rationality tries to avoid such problems by avoiding Dutch Book arguments in favor of an appeal to the requirements of consistency. The personalist Bayesians Colin Howson and Peter Urbach have argued for this approach. They wish to divorce the foundations of Bayesian probability theory from any consideration of the preferences of human agents. On their approach, probability theory should not be thought of as a theory of rational belief directly, but rather as a species of logic distinct from deductive logic. Just as deductive logic constrains the consistent assignment of truth values, "the logic of consistent assignments of *fair betting quotients* determined by appropriate constraints on them we can legitimately call *probability logic*" (Howson & Urbach, 2005, 66; emphases in original). They moreover propose that such a logic can be developed independently of considerations of expected gain or loss, if one assumes only the following principle:

Principle 1 *If the sum of finitely many . . . fair bets uniquely determines odds O in a bet on proposition a, then O are the fair odds on a.* (Howson & Urbach, 2005, 67)

In other words: adding together bets that are individually fair yields a compound bet that is also fair.

One might wonder how well the purely logical approach can succeed in excluding preference-based considerations while also relying on the notion of *fair* betting rates. The pragmatic approach understands a fair betting rate as a rate at which a person has no preference regarding which side of a hypothetical bet she would take. Howson and Urbach, however, propose to bypass this issue by identifying a person's fair betting rates instead with that person's *chance-based odds* regarding the proposition at issue. Chance-based odds are "*judgments about the relative likelihoods* of the relevant proposition and its negation" (Howson & Urbach, 2005, 53; emphasis in original). Although these are conceptually distinct from fair betting odds, they are numerically equal to them, because the betting rate which a person considers fair for a given proposition is that which the person considers to balance the risk between two sides of a bet. If Samiksha considers the chance-based odds of *Salt* being true to be 0.3, then that is also the betting rate at which neither side of a wager regarding *Salt* would enjoy an advantage over the other, according to Samiksha's estimation. What Samiksha would *prefer* is a separate question.

This logical approach to Bayesianism enjoys some advantages over the pragmatic approach. Critics have charged the latter with being both too psychological and anthropocentric to provide an adequate account of rational inductive inference. The psychological component of the pragmatic approach has been accused of relying on an unrealistic theory about human preferences and beliefs. By classifying Bayesian probability theory as a purely formal *logic*, Howson and Urbach seek to avoid these complications.

That avoidance comes with some costs, however. The pragmatic approach aspires to give an account of the *meaning* of probability statements in terms of the belief states of agents. But the logical approach effectively abandons this interpretive project, telling us only that a person's probability for a given proposition reflects that person's "chance-based odds" which are in turn "judgments about the relative likelihoods" of the proposition and its denial. But this tells us nothing about the *meaning* of statements about "chance-based odds," or what in general makes such statements true or false.

Howson and Urbach's probability logic also suffers from an important *weakness* of deductive logic. Because it is only concerned with consistency, deductive logic tells us only what conclusions follow from which premises (those whose denial would be inconsistent with the given premises), but tells us nothing about which premises we should begin with. Consequently, deductive logic can only tell us which combinations of beliefs we *might* hold consistently, but not which of those we *should* believe. Philosophers have often thought that inductive reasoning should fill this gap and supply us (fallibly, of course) with premises that we can then use in the context of deductive reasoning. But, by their own admission, Howson and Urbach's probability logic can no more do this than can deductive logic. It tells us what combinations of probabilities may consistently be held collectively, but not which of the infinitely many such combinations for any given set of propositions is the right probability assignment for that set. They regard that task as unachievable by any account of inductive reasoning.

8.5 Bayes's theorem and updating

We have thus far been focusing on the foundations of the Bayesian account, but have not yet gotten to Bayes's theorem, the powerful result from which Bayesianism derives its name.

Before stating the theorem, let us note that in our concern with foundations, we have been dealing only with unconditional probabilities like $Pr(Salt)$. What about conditional probabilities, like 'the probability that Tatyana can eat 10 saltines in 60 seconds without drinking anything, *given that* Tatyana is a world champion in competitive eating'? Without going into details, we can simply note that we can extend the wagering analysis we have employed previously quite naturally with the concept of a 'conditional bet,' in which the parties agree to a wager at a certain rate, but on the condition that a certain proposition is true. If that proposition turns out not to be true, then there is no wager.

Bayes's theorem concerns conditional probabilities. Bayesians contend that the axioms of probability theory constrain what combinations of beliefs may consistently be held at a given time (called *synchronic* constraints). But to give an inductive account of how science progresses, or more generally how people learn, we need a *diachronic* perspective that can handle the situation where a person receives new information. How should beliefs *change*?

The Bayesian approach to this supposes that a rational agent starts with a certain set of degrees of belief that can be represented with a *probability function*, i.e., for every proposition p that is a subject of belief for an agent S, there is a probability that represents S's degree of belief in p. We will call that initial probability function Pr. Suppose that we are concerned with S's belief regarding some particular claim H and we are interested in how that belief should change upon learning of the truth of some proposition E (in this standard notation, H can be thought to suggest 'hypothesis' and E, 'evidence'). To answer that question, we need some additional information: (1) What is the probability that E is true, supposing that H is true? and (2) What is the probability that E is true, independently of the truth of H? (The latter is sometimes called the *expectedness* of E.) Given this information, we can apply Bayes's theorem, which in its simplest form can be stated using the equation:

Theorem 1 (Bayes's theorem)

$$Pr(H|E) = \frac{Pr(E|H)Pr(H)}{Pr(E)},$$

where Pr is a probability function and $Pr(E) \neq 0$. The quantity on the left side of Theorem 1 is the *posterior probability* of H, given E, in contrast with $Pr(H)$, the *prior probability*. Note that the posterior probability is a conditional probability, and for personalist probabilities Bayes's theorem relates this

quantity (an agent's fair betting rate for a wager on H, with the bet to proceed only on condition that E is true) to an agent's probability for H independently of E, her probability for E independently of H, and her probability for E conditional on H.

To address the question about how beliefs should change, we need to add a rule that relates this conditional probability $Pr(H|E)$ to the unconditional probability that the agent should attribute to H once she *has* learned that E is true. For the Bayesian, change in belief is a matter of replacing one probability function with another one. Rules for doing so are called *updating* rules, and Bayesians rely on *updating by conditionalization*.

Principle 2 (Updating by conditionalization) *At time $t = t_0$, let S's state of belief be represented by a probability function Pr. If, at $t = t_1 > t_0$, S acquires the information that E is true (and no other new information is acquired), then at t_1 S should adopt the probability function Pr', such that for every proposition ϕ that is a subject of belief for S, $Pr'(\phi) = Pr(\phi|E)$.*

Of course, $Pr(\phi|E)$ is precisely the posterior probability of ϕ conditional on E that we can calculate using Bayes's theorem. Although updating by conditionalization is central to the Bayesian approach, it cannot be derived from the probability axioms via any 'Dutch Book' argument (though see Teller, 1973). Howson and Urbach conclude that the rule is valid only for an agent who leaves her conditional probabilities unchanged when she acquires additional information E. In other words, $Pr'(\phi) = Pr(\phi|E)$ provided $Pr'(\phi|E) = Pr(\phi|E)$. But the probability axioms do not require an agent to keep the same conditional probabilities over time.

Nonetheless, assuming that we may attribute degrees of belief to agents engaged in inquiry under conditions of uncertainty, and assuming those agents maintain stable conditional probabilities, the Bayesian now has a theory of how such agents must revise their beliefs, if they are to remain coherent.

8.5.1 Example: fossil origins

To see how this might illuminate scientific reasoning, consider an artificial, though historically inspired, example.[7] For a long time *fossils* posed a great

[7] The example that follows draws heavily on Martin Rudwick's authoritative study (1985), which was brought to my attention by Kyle Stanford. Stanford's own (2010) offers a

mystery. They *looked* somewhat like organisms,[8] but they were not made of organic material; they were found inside of rocks, in places far from where one would expect to find organisms of the type they resembled (such as marine-looking forms found in the Alps); many of them corresponded in form to no living organisms. From the Renaissance, when serious study of the issue seems to have begun, until well into the eighteenth century, the idea, so familiar to us today, that they resulted from the death and fossilization of organisms, was at best a problematic proposal, and for most of that time was justly regarded as less convincing than competing explanations.

To simplify greatly, we will pretend that there were just two such competitors to the hypothesis of organic origins: Renaissance Neoplatonists believed that all of nature was related by a network of affinities and correspondences, rooted in an "ontological analogy between Man [microcosm] and his external world [macrocosm]," "which might be made manifest by resemblances not only between microcosm and macrocosm, but also between the heavens and the Earth, between animals and plants, and between living and non-living entities" (Rudwick, 1985, 18–19). Neoplatonists considered the creation of organic-looking forms in inorganic material the work of "the very same 'moulding force' or *vis plastica*" that directed "the growth and development of living organisms themselves" (Stanford, 2010, 222). Aristotelians developed an alternative approach. Since they held that spontaneous generation could give rise to living organisms when a living form acted on non-living matter, it was natural to suppose that similar structures could be generated in rocks "by the form present in their characteristic seeds acting on the materials of the Earth, carried there by such natural processes as the percolation of groundwater" (Stanford, 2010, 223; see Rudwick, 1985, 34–35).

These two views we shall attempt to codify with the hypotheses *NP*: 'Fossils are formed through the action of the *vis plastica* on inorganic materials to manifest their affinity with organisms' and *A*: 'Fossils are formed by seeds of living things imparting their forms to inorganic materials.' Finally, we have the organic origins hypothesis *O*: 'Fossils are formed by processes that transform the remains of organisms into mineralized forms.'

fascinating (non-Bayesian) discussion of the changing nature of the evidence for the organic origins of fossils from the period discussed here to contemporary times.

[8] Some of them did, at any rate. The term 'fossil' was for centuries used "to describe *any* distinctive objects or materials dug up from the earth or found lying on the surface" (Rudwick, 1985; emphasis in original).

Such views were well known to Niels Stensen (1638–1686) in 1665, by which time he had arrived, a renowned anatomist, in Florence from his native Copenhagen, Denmark, via Leiden and Paris. (The anglicized version of his name 'Steno' appears in most scholarly literature.) While in Florence, Steno undertook to compare *glossopetrae* ('tongue-stones') with the teeth of living (more precisely, until-quite-recently-living) sharks, which tongue-stones had already been noted to resemble, and found support for the hypothesis of organic origins. Steno's observations identify two salient facts beyond the mere fact that such fossils resemble sharks' teeth with regard to their shape (Neoplatonists and Aristotelians alike claimed the ability to account for this): (1) tongue-stones show signs of *decay* and (2) tongue-stones show no signs of distortion of the sort seen in tree roots when they grow in rock crevices, but are "the same shape, whether they come to light from softer ground, or are chipped out of rocks" (Steno, [1667]1958, 15). The first point is significant, Steno argues, because it implies "that they were not being formed at the present time but were relics of an earlier period." The second fact indicates that "the surrounding 'earth' must have been soft when it first enclosed them" (Rudwick, 1985, 50).

Steno's argument seems to be this: According to both Neoplatonist and Aristotelian theories, the processes that produce fossils act on a relatively short time scale (the time scale of the growth of individual living organisms) and are not restricted to any particular period of history. We should not expect to find indications of decay in all such tongue-stones, then, but should find them to be intact, just as in the case of recently formed teeth in living sharks. Moreover, both Neoplatonist and Aristotelian theories suppose that fossils grow inside rocks. But where we see other examples of things growing in rocks (as in the case of tree roots), we find that the difficulty of displacing the surrounding rock causes the form of the growth to be contorted. Neither of these conclusions is compatible with Steno's observations, so he concludes that these facts all point "unanimously" to the conclusion that tongue-stones formed where sharks had died, followed by a process of sedimentation during which the surrounding rocks formed.

Bayesians have a fairly straightforward way to understand these arguments (with a little bit of reconstruction): Steno is arguing that these observed characteristics of tongue-stones are of a sort that hypothesis *O* renders probable. We should expect to see just such traits, if *O* is true. The hypotheses *A* and *NP*, by contrast, make it improbable that we would observe

such facts. Moreover, the plausibility of *O* prior to the consideration of these facts is not insignificant, while it is not highly probable that there is some other hypothesis that no one has thought of yet that would make sense of these observations.

The argument thus characterized in qualitative terms can be connected more strongly with the Bayesian apparatus if we agree to engage in a little historical fiction and attach numbers to these judgments. The numbers are not likely to represent any of Steno's actual beliefs. However, if we grant the Bayesian claim that beliefs formed under conditions of uncertainty can be represented as susceptible to degrees, then, at least for the purposes of illustration, we might postulate a fictional, Steno-like character – let's call him 'Schmeno' – whom we assume to be coherent in the Bayesian sense and whose degrees of belief the probabilities given below might approximate. We will use *S* to refer to Schmeno, *E* to represent the proposition 'A collection of tongue-stones was examined and found to show signs of decay and not to show signs of distortion' and *CA* to represent the catch-all hypothesis 'Fossils form as the result of some process other than those referred to in *NP*, *A*, and *O*.'

We will start with the prior probabilities:

$$\begin{array}{rclrcl} Pr_S(NP) & = & 0.25 & Pr_S(A) & = & 0.25 \\ Pr_S(O) & = & 0.25 & Pr_S(CA) & = & 0.25 \end{array}$$

In effect, we are supposing that Schmeno begins in a state of indifference with regard to the hypotheses of Neoplatonism, Aristotelianism, or organic origins as regards the explanation of fossils. He is also indifferent, relative to each of those hypotheses, with regard to the possibility that there is some other explanation of which he remains unaware.

Now for the probability of the evidence under these various hypotheses:

$$\begin{array}{rclrcl} Pr_S(E|NP) & = & 0.1 & Pr_S(E|A) & = & 0.1 \\ Pr_S(E|O) & = & 0.8 & Pr_S(E|CA) & = & 0.5 \end{array}$$

Here we are supposing that Schmeno regards the observations that he has made as rather improbable if either *NP* or *A* is the correct explanation for fossils, since under both of those hypotheses, he believes, he probably would not find in all his samples signs of decay and the absence of contortions. But the organic origin hypothesis makes such observations probable, in the sense that one should expect to find them. (Note that the purely fictional

value $Pr_S(E|O) = 0.8$ simply expresses the vague assumption that Schmeno considers his observations to be pretty much expected but not guaranteed by the organic origins hypothesis. In those cases where a prediction can be *deduced* from a hypothesis, then the probability of the prediction, conditional on that hypothesis is equal to one. In many other cases, the probability of a particular observation can be calculated from a quantitatively expressed hypothesis, perhaps with the help of auxiliary assumptions.) Meanwhile, if some other explanation of these observations of fossils is correct (i.e., the catch-all hypothesis is true), then it is unclear whether one should expect these observations or not. Schmeno considers E in that case to be an even-odds proposition.

Updating by Bayesian conditionalization will lead Schmeno to a set of posterior probabilities. But it will help here if we use a slightly different form of Bayes's theorem. In the following we take advantage of the fact that, if $H_1, H_2, \ldots, H_n$ constitute a set of mutually exclusive propositions such that for some probability function Pr, $Pr(H_1 \vee H_2 \vee \cdots \vee H_n) = 1$ (i.e., with probability one, at least one of the propositions $H_1, H_2, \ldots, H_n$ is true), then for any proposition Q,

Theorem 2

$$Pr(Q) = Pr(Q|H_1)Pr(H_1) + Pr(Q|H_2)Pr(H_2) + \cdots + Pr(Q|H_n)Pr(H_n).$$

We apply Theorem 2 (sometimes called the *total probability theorem*) to the denominator in the right-hand side of Bayes's theorem (Theorem 1) to get the following:

Theorem 3 (Bayes's theorem expanded)

$$Pr(H|E) = \frac{Pr(E|H)Pr(H)}{Pr(E|H_1)Pr(H_1) + Pr(E|H_2)Pr(H_2) + \cdots + Pr(E|H_n)Pr(H_n)},$$

where $H_1, H_2, \ldots, H_n$ constitute a set of mutually exclusive propositions such that $Pr(H_1 \vee H_2 \vee \cdots \vee H_n) = 1$ (i.e., the Hs exhaust the possibilities) and $Pr(E|H_i)Pr(H_i) > 0$ for some $1 \leq i \leq n$ (i.e., the denominator is not equal to zero).

This form of Bayes's theorem provides some guidance for estimating the expectedness of E, which otherwise seems inscrutable. The prior probabilities of the hypotheses and the probabilities of E under those hypotheses provide us the means to determine the expectedness of E.

Applying Theorem 3 to Schmeno's probabilities, then, we arrive at the following posterior probabilities:

$$\begin{aligned} Pr_S(NP|E) &= 0.07 \quad Pr_S(A|E) = 0.07 \\ Pr_S(O|E) &= 0.53 \quad Pr_S(CA|E) = 0.33 \end{aligned}$$

We can see here how the probabilties have been redistributed in light of E. The probability of O has increased significantly, to a point where Schmeno regards it as more probable than not. Meanwhile, NP and A have significantly decreased in probability. The probability of CA has also increased slightly.

Of course, were Schmeno to begin with different prior probabilities, then his posterior probabilities would also differ, and nothing in the Bayesian account constrains prior probabilities beyond the requirement that they be coherent.

Although this may appear to be a problem, Bayesians consider it a strength of their account. Recall how Kuhn emphasized that an account of the development of scientific thought should allow scientists to rationally disagree. Allowing coherent Bayesian agents to have different prior probabilities accomplishes that.

For example, the naturalist Martin Lister (1638–1712) disagreed with Steno's conclusions. When confronted with Steno's arguments, Lister found them unpersuasive not because he disagreed with the observations Steno had made or their relevance for the hypothesis in question, but because he found the hypothesis of organic origins for fossils highly implausible on the basis of other considerations drawn from his own study of shell-like fossils: (1) the substance of these shell-like fossils did not include any actual shell; (2) fossils were found only in certain kinds of rock formation, suggesting that they preferentially grow in those kinds of formations in the same way that plants grow preferentially in certain kinds of soils; (3) although such fossils resemble shells, they do not resemble the shells of any existing species (the idea of widespread extinctions of species would have been entirely foreign to Lister). So Lister rejected the general claim that fossils could be understood as the result of some process whereby organic remains retained their shape but became "petrified" (Rudwick, 1985, 62–63).

We can avail ourselves again of the Bayesian framework to see how Lister, while not rejecting Steno's observations, might reject his conclusions. Again, in the absence of sufficient historical accuracy with regard to Lister's actual beliefs, we will have to invent a Lister-ish character to whom we may attribute

a set of Bayesian probabilities. Let's call him 'Listerio.' We will add L to our existing notation to refer to him.

We can begin by stipulating that Listerio and Schmeno are in complete agreement over the probabilities imparted to Schmeno's observations by the hypotheses under consideration:

$$\begin{array}{rclrcl} Pr_L(E|NP) & = & 0.1 & Pr_L(E|A) & = & 0.1 \\ Pr_L(E|O) & = & 0.8 & Pr_L(E|CA) & = & 0.5 \end{array}$$

But to reflect Listerio's skepticism about the hypothesis *O*, we should attribute a different set of prior probabilities. Whatever Lister may have made of the Neoplatonist and Aristotelian viewpoints, we can pretend that Listerio found them neither more nor less plausible than did Schmeno. If he is to begin with a lower prior probability for *O* than Schmeno, he will need to attribute a greater probability to the catch-all *CA*. So we will stipulate the following priors:

$$\begin{array}{rclrcl} Pr_L(NP) & = & 0.25 & Pr_L(A) & = & 0.25 \\ Pr_L(O) & = & 0.05 & Pr_L(CA) & = & 0.45 \end{array}$$

This leads to the following posterior probabilities:

$$\begin{array}{rclrcl} Pr_L(NP|E) & = & 0.08 & Pr_L(A|E) & = & 0.08 \\ Pr_L(O|E) & = & 0.13 & Pr_L(CA|E) & = & 0.71 \end{array}$$

Note here that Listerio, like Schmeno, judges Schmeno's observations to raise the probability of *O*, and to reduce the probabilities of *NP* and *A*. Those relative changes are forced by the probabilities conferred on those observations by the hypotheses, on which Schmeno and Listerio agree. But for Listerio, the posterior probability of *O* remains rather low, while the probability of the catch-all has increased significantly.

There are two important points to emphasize in our fake historical example. The first is that, by Bayesian standards, Schmeno and Listerio are both (according to our stipulations) being perfectly rational, in spite of their disagreement. The second point is that Lister based his rejection of Steno's argument on *his observations* of shell-like fossils found in England. What we have depicted as a low prior probability for *O* could also be regarded as a low *posterior probability* relative to Listerio's observations. As deployed by personalist Bayesians, the notions of prior and posterior probability are always relative to some particular set of data. (The flip side of this observation is that

Bayesian updating by conditionalization does not apply to the problem of how to redistribute probabilities in the face of a newly invented hypothesis.)

8.6 Problems for Bayesianism

8.6.1 Subjectivity and the catch-all hypothesis

I have been rather glib about the phoniness of this reconstruction I have given. Certainly one could more conscientiously take historical figures' actually expressed beliefs into account when attempting a Bayesian reconstruction of some historical episode.[9] The point of the illustration is not the actual numbers we arrive at, but the positive or negative *effect* that evidence has on the probability of a given hypothesis, and how that effect compares to what happens with regard to competing hypotheses, for agents starting from different prior probabilities.

Nonetheless, a little reflection on this reconstruction highlights a difficulty for Bayesianism entirely independent of any problems we might have coming up with realistic probability assignments to reflect scientists' beliefs: the dependence of the Bayesian calculation on the catch-all hypothesis.

Critics of Bayesianism often point to the fact that it does not dictate prior probabilities, leaving them as a 'subjective' matter. Subjectivity, such critics argue, should be avoided in a theory that purports to tell us how we should draw conclusions from scientific data. John Norton sees the problem as inherent in the personalist Bayesian shift that merges the notions of 'degree of support' and 'degree of belief' into a single representation in terms of probabilities. The result of this merger is that "[t]he evidential relations that interest us are obscured by a fog of personal opinion" (Norton, 2010, 507).[10]

We have already noted the standard Bayesian replies to criticisms of the subjectivity of prior probabilities: First, most prior probabilities are posterior to some previously considered evidence. Second, it is only sensible that

[9] See (Dorling, 1979) for an example of this that seeks to use Bayesianism to address the underdetermination problem, and see (Mayo, 1997a) for a critique.

[10] The focus of Norton's argument against Bayesianism is that it fails to distinguish between cases where the evidence *fails to support* some hypothesis (such as when evidence is completely neutral with regard to a class of competing hypotheses) from cases where the evidence *supports the denial* of a hypothesis. This conflation results, he argues, in "spurious" Bayesian inferences.

plausibility judgments should make a difference to how scientists evaluate the evidence supporting theories; this just clarifies in what sense scientists can rationally disagree about the status of a theory. The catch-all points to a subjectivity problem that goes deeper than these responses acknowledge.

Catch-all hypotheses introduce subjectivity into Bayesian inference in two ways. Recall that in order to calculate the denominator when applying Theorem 3 to our example, we needed *both* $Pr_S(CA)$ and $Pr_S(E|CA)$. The former, of course, is a prior probability and hence subjective to the extent that any prior probability is. Estimating the latter quantity, the probability of the evidence under the hypothesis *CA*, seems an elusive task. In general, probabilities of particular data or observations conditional on hypotheses under consideration are treated by Bayesians as unproblematic and agreed upon by scientists, because they can often be calculated from the hypotheses in question (perhaps with the help of auxiliary hypotheses). But asking for the probability of the data under the catch-all seems to ask us to do the impossible: to determine the probability of some specific data or observation, conditional on a completely unspecified potentiality: '*some other hypothesis is correct*'.

In our example, we specified $Pr_S(E|CA) = 0.5$, but why? That assignment seems to reflect a judgment that, if none of the hypotheses considered is correct, then *E* is just as likely to be true as not. But that judgment has no justification. Indeed, it is hard to imagine what *would* justify anyone in assuming this or any other value for the quantity at issue. If Schmeno had a belief such that $Pr_S(E|CA) = 0.9$, there would be no Bayesian grounds for criticizing him, nor if he held beliefs such that $Pr_S(E|CA) = 0.1$. If the former were true his posterior probability for *O* would be only 0.42, which would be less than the posterior for *CA*, and if the latter were the case, then the posterior for *O* would be 0.73. The probability of *E* under the catch-all can thus have a significant influence on the posterior probability.

8.6.2 Relevance and normativity

Even if we grant, however, that the Bayesian account allows for the articulation of a probabilistic theory of rationality, it remains somewhat unclear how the notion of rationality that emerges relates to the aims of scientific inquiry. For example, we might think that the point of scientific inquiry is to discover theories that are true, or approximately true. Or, we might consider the point to be the discovery of theories that are generally reliable. According

to the pragmatic approach to the Dutch Book argument, the person who holds incoherent degrees of belief is irrational in the sense of being open to a combination of bets that guarantee a loss. How is this relevant to the scientist who seeks to assess the merits of a scientific theory in the face of a body of evidence? The scientist cannot be regarded as even implicitly making a wager on the truth, approximate truth, or general reliability of a general theory, but at most on that theory's *reliability* in some particular series of applications. This is because any agreement to a betting contract (even a hypothetical one) must include agreement to a criterion by which to settle the bet in some finite time. But the truth, approximate truth, and general reliability of theories are all open-ended in the sense that any judgment regarding them is provisional. If we regard any of these as the aim of scientific inquiry, then scientists' preferences regarding wagers are irrelevant to the aims of science, which concern matters not subject to wagering.

A further problem for Bayesians concerns the very nature of their project. One aspect of the logical empiricist project that both Bayesian and non-Bayesian methodologists have inherited is an interest in developing accounts of scientific reasoning that not only describe how scientists reason but that are *normative* in the sense that they can provide guidance to investigators seeking to reason *well*. This normative aspect of the Bayesian project is evident from their focus on rationality. Yet the Bayesian notion of rationality is arguably inappropriate for the purposes of a normative philosophy of science, being either so weak as to be trivially satisfied or so demanding as to prevent its being applicable to the situations encountered in practice by scientists.

Consider the situation of a scientist S who, acquainted with certain phenomena E in a given domain, entertains some number of competing hypotheses $H_1, \ldots, H_n$ as potential explanations of the phenomena in that domain. S wishes to know what inferences regarding those hypotheses might be warranted. The first weakness of Bayesianism arises at just this point: S might think that the situation calls for the collection of further data, perhaps disclosing new phenomena, to test these hypotheses. The Bayesian approach does not obviously provide guidance as to what kinds of procedures S ought to employ to produce data that will help to discriminate amongst those hypotheses. It only tells S that the data will not make a difference to her prior rankings of these hypotheses if they are equally probable under each of them. That is not much help.

The Bayesian might counter that giving more concrete methodological guidance is not the task of a theory of rationality. The theory need only allow for the discrimination between beliefs that are jointly held reasonably and beliefs that are not thus held. So let us consider the question of whether S is rational to hold a certain set of degrees of belief regarding $H_1, \ldots, H_n$, in light of E. According to the Bayesian approach, we should think of these degrees of belief as the result of an updating, in which S, having gained knowledge of E, replaced one belief function with another belief function. Then this belief function would be rational in the Bayesian sense if that updating were to be carried out by applying the rule of conditionalization, i.e., the new degrees of belief are equal to posterior probabilities calculated by applying Bayes's theorem to a prior belief function, where that prior belief function was coherent (i.e., satisfied the probability axioms).

In one sense, this requirement is very *easy* to satisfy. Whatever one's current degrees of belief, so long as they are presently coherent, there will be *some* prior probability function that, by Bayesian updating, would yield precisely that posterior probability function. Thus it seems that if we are committed to viewing S's updated beliefs as rational, then (assuming S is presently coherent) we can vindicate that judgment by attributing to S the requisite prior beliefs, which, after all, are not directly observable. The apparent triviality of its requirements renders doubtful the value of Bayesianism even for the limited task of rationally reconstructing historical episodes (Mayo, 1997a).

Note, however, that this argument did require that S's present beliefs must be synchronically rational, meaning that they must be coherent in the Bayesian sense. As long as we confine our attention to a small number of propositions, it is easy enough to come up with a plausible probability assignment that meets these requirements. But Bayesian rationality requires that *all* of S's degrees of belief must jointly satisfy the probability axioms. Because of the large number of beliefs that one must hold at any given time and the subtle ways in which different beliefs are related to one another, giving a consistent explicit assignment of subjective probabilities to some small number of hypotheses under consideration might nonetheless yield incoherence in one's overall set of beliefs (Senn, 2001, 2011). Thus, even when a reconstruction appears to show the rationality of a set of beliefs in the face of some body of evidence, a real person holding such beliefs might be incoherent because of other, off-stage beliefs.

8.7 Bayesian philosophy and Bayesian statistics

Our discussion thus far has considered Bayesianism as a theory of rational belief under conditions of uncertainty, to be used to examine real or hypothetical episodes of scientific reasoning and, through a process of reconstructing the inference in Bayesian terms, display its rationality or lack thereof. Can Bayesianism be used by scientists to help them answer questions?

For many years, philosophical regard for Bayesian theory far exceeded the respect afforded it by statisticians, who found it computationally difficult to apply to the data analysis problems they faced. Applied statistics was dominated by methods rooted in the frequentist approach to probability that the next chapter discusses, while Bayesianism has dominated philosophy of science (and more recently has percolated into other areas of philosophy such as epistemology and philosophy of religion).

While much statistical analysis of data across the sciences remains frequentist in orientation, the field of *Bayesian statistics* is growing. Although there is no rigorous definition of the term, statisticians tend to regard an analysis of data as Bayesian just in case it makes use of a prior probability distribution over the range of hypotheses of interest. Bayesian statistical analyses need not be rooted very deeply in the personalist Bayesian philosophical theory, in that the prior probabilities they use may not be intended to reflect some person's pre-data state of opinion regarding the hypotheses at stake. Bayesian statisticians often choose instead to use a *non-informative* prior distribution that distributes probability evenly among all competing hypotheses (Jaynes, 2003). Another approach advocates the use of *reference priors*. These are prior probability distributions chosen so that the distance (according to a particular measure) between the posterior distribution and prior distribution will be maximized (Bernardo, 1979; Berger & Bernardo, 1992). Many Bayesian statistical analyses concern problems in which there is a prior assumption that a certain parameter has a numerical value, possibly within some determinate range, and the hypotheses under consideration concern the value of that parameter. Provided that the prior assumption is correct (often not a trivial issue), the bite of the worry previously raised about the catch-all hypothesis is considerably mitigated, since all of the hypotheses consistent with the assumption can be stated explicitly and often yield determinate probabilities for the data under consideration.

The idea that the usefulness of Bayesian statistical methods is entirely independent of the status of Bayesian philosophy has a pragmatic appeal. Many scientists who have used them value the flexibility of Bayesian methods. But philosophical issues, particularly concerning *interpretation*, lurk wherever we find statistics used. Regardless of how the prior distribution is chosen and justified, the point of a Bayesian analysis is to generate a *posterior* distribution, which serves as the output of the analysis. For a personalist Bayesian, the meaning of this posterior distribution is clear: it represents the state of belief that a rational agent should have with regard to the relevant hypotheses, in light of the data, if that agent starts out with a state of belief represented by the prior distribution. If the prior distribution does not represent a state of belief, but is chosen because it is non-informative, or affords ease of computation, or for some other practical reason, then the posterior distribution no longer has this same meaning. But then what does it mean? David Cox and Deborah Mayo have stressed the need for objective Bayesian statistics to clarify how its posterior probabilities fulfill the goal of measuring the strength of the evidence supporting a given inference (Cox & Mayo, 2010).[11]

8.8 Conclusion

Does Bayesianism, in the personalist form that has been the primary focus here, provide the kind of neutral framework for the evaluation of evidence that Carnap and other logical empiricists sought? In one sense it does not. Carnap sought a measure of the degree to which any given data confirm any given hypothesis that was *objective*, in the sense that any two inquirers who agreed on the language to be used to describe the hypothesis and the data would also agree on the degree of confirmation. But personalist Bayesian inquirers can disagree on this, if they start out with different prior probabilities or if they disagree on how probable the data are given the competing hypotheses.

In another sense, however, the Bayesian framework does exhibit some features of the logical empiricists' vision. Although disagreements over prior probabilities can yield discrepant posterior distributions, Bayesian

[11] In an interesting effort to address such foundational issues, Jon Williamson has pursued an interesting variant of Carnap's objective version of Bayesian inference (Williamson, 2010).

investigators will all share the same procedure for updating from the former to the latter. In other words, they will agree on how the data are relevant to the evaluation of the hypotheses at issue. Moreover, as Bayesians stress, if agents with differing prior probabilities continue to update by conditionalizing on the same accumulating body of data (and they agree on the probability of the data, and they agree about which hypotheses have probabilities of 0 or 1), their repeated updating will yield a *convergence* of their probability functions (see Doob, 1971; Gaifman & Snir, 1982; Earman, 1992, 144–149 includes a philosophical assessment). Thus, although Bayesians attribute subjectivity to scientific reasoning, they provide a framework for reducing that subjectivity.

Noting these points, Wesley Salmon proposes that Bayes's theorem provides a means for building a bridge between the logical empiricist project and Kuhn's post-positivist insights (Salmon, 1990). Salmon argues that we can interpret the kinds of paradigm-dependent judgments that lead scientists to evaluate theories differently in Kuhn's account as yielding different prior probabilities representing plausibility judgments before some body of data had been taken into account, while Bayes's theorem provides a paradigm-independent device for bringing those data to bear on the theories in question. Nonetheless, Salmon worries that Bayesianism in its usual form allows too much subjectivity into scientific reasoning, because of the difficulty of determining the probability of the evidence under the catch-all hypothesis. Salmon proposes a solution that uses Bayes's theorem, not to determine a posterior probability as Bayesians do, but to *compare* for any given pair of hypotheses, which of them ought to be preferred over the other, given the data in hand. His approach eliminates the need to refer to the catch-all hypothesis, but at the cost of greatly limiting the strength of one's conclusions: Salmon's approach leads to conclusions only about which of the theories under consideration we ought to prefer to the others under consideration, but not to conclusions about the truth, probability, or truthlikeness of those hypotheses.

Debates between defenders and critics of Bayesianism are among the most hotly contested in philosophy of science as well as in statistics, where the arguments at one time became so pugilistic that they became known as the 'statistics wars.' Here we have barely scratched the surface of the issues. A historically significant collection from the statistics wars is (Savage, 1962). Howson and Urbach's (2005), on which I have drawn for the exposition here,

gives a good overview of many of the criticisms of Bayesianism in the context of a vigorous defense of the view. Deborah Mayo, whose frequentist-based approach to inference we discuss in the next chapter, has voiced some of the most significant criticisms of Bayesianism (Mayo, 1996). Earman (1992) gives a mixed assessment of Bayesianism, surveying a number of debates that continue to develop.

8.9 Appendix: solution to the Chevalier's problem

Turning first to Game 2: That the die lands 6 up at least once in four tosses is equivalent to the denial of the claim that it does not land 6 up in any of four tosses. The latter claim is equivalent to the statement that the die lands with something other than 6 up on each of four tosses. To calculate the probability of Game 2 we can therefore calculate the probability of getting not-6 four times in a row, and then subtract that number from 1. (Recall that the probability of a logical truth such as 'p or not-p' is (by Axiom 1) equal to 1, so for any proposition p, $Pr(p) = 1 - Pr(\textit{not-}p)$.) The probability of getting something other than 6 on a single toss is 5/6, and that probability for any given toss is independent of what happens in any other toss. So (applying the definition of conditional probability) the probability of not-6 on four consecutive tosses is $5/6 \times 5/6 \times 5/6 \times 5/6 \approx 0.482$. The probability of getting at least one 6 is then approximately $1 - 0.482 = 0.518$.

Calculating the probability of winning Game 3 uses a similar technique. There are 36 possible, equally probable outcomes for a toss of two dice, only one of which consists of two 6s. The probability of *not* getting two 6s on a single toss is therefore 35/36, and on 24 tosses is $(35/36)^{24} \approx 0.509$. The probability of getting a pair of dice to land 6 up at least once in 24 tosses is therefore approximately $1 - 0.509 = 0.491$.

9 Reasoning with probability: frequentism

9.1 Introduction

In the last chapter, we explored that aspect of probability ideas concerned with "the degree of belief warranted by evidence." Now we turn to the other aspect: "the tendency, displayed by some chance devices, to produce stable relative frequencies" (Hacking, 1975, 1). Our use of examples involving games of chance to introduce the mathematics of probability at the beginning of the previous chapter relied implicitly on this idea. But are dice games relevant to science? As we will see, thinking of *scientific experiments* or *measuring procedures* as chance mechanisms with tendencies "to produce stable relative frequencies" holds the key to the potential value of relative frequency ideas for philosophy of science.

Frequentism understands probability statements as statements about the relative frequency with which a certain outcome would occur under repeated execution of some process. Frequentist ideas have been central to the development of theoretical and applied statistics, and thus have had a significant influence on how scientists analyze data and report their results. Some of these statistical practices have been subjected to considerable criticism, however, and philosophers of science have tended to regard Bayesianism as based on a more coherent set of principles. Recently, the *error-statistical philosophy* has attempted to respond to philosophical criticisms of frequentist statistics, while remaining founded on the idea of experiments as chance mechanisms.

Since, however, error statistics arises from an attempt to reform and rehabilitate the frequentist tradition in statistics, we will start with a brief introduction to that tradition. Statisticians have long emphasized the importance of experience in the use of statistical methods and analysis of real data to a good understanding of statistics, and with good reason. What follows is too brief and superficial to underwrite a useable understanding of statistics.

Nonetheless, we may hope that some grasp of the conceptual underpinnings of frequentist statistics will convey a sense of the philosophical orientation and import of these ideas.

9.2 Relative frequency 'in the long run'

Suppose that, following the suggestion floated early in the last chapter, we interpret the statement 'the probability of winning Game 1 is 1/6' to mean 'if you play Game 1 over and over again, then, on average, you are going to win one out of six games.' Although this interpretation has an intuitive appeal, we have taken an apparently definite mathematical statement and given it a vague and potentially misleading interpretation. It is vague because "over and over again" is vague. How many times is that? It is potentially misleading because if we think that playing over and over again is something that can be achieved by playing some finite number of times, then the statement 'if you play Game 1 over and over again, then, on average, you are going to win one out of six games' might not be *true*, even if 'the probability of winning Game 1 is 1/6' is true. As we will see, you might actually win either more or less than that. (The name we have for that phenomenon is 'luck'.) Let's see if we can do better.

As Hans Reichenbach notes, the frequency interpretation of probability has two distinct functions in probability theory. One is concerned with providing an account of the *meaning* of probability statements, while the other is concerned with accounting for the *substantiation*, or support, of probability statements (Reichenbach, 1938, 339). We will begin with the first function.

Suppose that you do undertake to play Game 1 repeatedly, recording after each roll of the die the number of times you have played N and the number of times you have won N_w. I did this.[1] Table 9.1 gives my recorded values for the fraction N_w/N – the *relative frequency* of winning – after 12 rolls.

Such bad luck! I continued rolling the die,[2] and after 36 rolls, the relative frequency of wins was still only $N_w/N = 1/36$. At $N = 216$ I had $N_w/N = 23/216 \approx 0.11$ (note that $1/6 \approx 0.17$). At $N = 3600$ (surely enough to

[1] No, I didn't. Being too lazy to sit and roll a die over and over again, and lacking the nerve to ask for funding to pay a student to roll a die over and over again, I used a random number generator that produced outputs of whole numbers from 1 to 6 to simulate my data. No conclusion about any actual die can be drawn from my fake data.

[2] Not really. See above.

Table 9.1 *Data from the rolling of a die.*

Roll	1	2	3	4	5	6	7	8	9	10	11	12
Result	5	6	1	3	2	5	6	6	6	5	3	6
N_w/N	0	0	0	0	0	0	0	0	0	0	0	0

make the point, so that I could take a break from the tedium of repeatedly rolling that die![3]), the relative frequency of wins was looking much more like the 'right' number: $N_w/N = 613/3600 \approx 0.17$.

In our example of the dice game, the relative frequency approach to probability interprets $Pr_M(X_1 = 4) = 1/6$ as a statement not (directly) about what happens in any finite sequence of rolls of the die, but about the value of N_w/N as N becomes infinitely large, or *in the limit as* $N \to \infty$. More generally, according to the frequentist view:

Definition 3 (Frequentist probability) *If one is able to repeat a trial under identical conditions and note whether some event E occurs as an outcome of the trial, and if* N_E/N *is the relative frequency with which E occurs, then*

$$Pr(E) \equiv \lim_{N\to\infty} \frac{N_E}{N} \tag{9.1}$$

Of course, this simple definition only applies if the probability that a certain kind of trial has a certain kind of outcome is the same from one trial to the next and is independent of how many trials have been carried out. Even repeatedly throwing a die would seem not to *exactly* satisfy this requirement: as we extend the number of throws to infinity, we will at some point undoubtedly notice the die start to show signs of wear, however long that might take, and the relative frequency of 4s might not remain the same as the slow process of erosion of the die continues. To suppose that it will remain the same constitutes an *idealization* in our probability model of Game 1. To be an appropriate idealization, it needs to be the case that although it may not be strictly true that $\lim_{N\to\infty} \frac{N_w}{N} = 1/6$, we will not be misled in any of our real-world reasoning if we pretend that it is.

The die-rolling data exhibits a pattern that, intuitively, seems typical: Although small amounts of data show a behavior that we might not have been able to predict, the more data we collect, the more the data align

[3] If I had actually been rolling it in the first place …

with what we think *ought* to happen (based on our theory about the die). The mathematical elaboration of frequentist probabilities allows us to make precise our intuitive sense of what is *typical*. Probability distributions are crucial to this clarification. They convey mathematically both what outcome is most to be *expected* and the degree to which the outcomes of a given process are *variable*.

To see how, consider the 3600 throws of the die to constitute a series of 100 *experiments*, in each of which the die has been thrown 36 times. In any given experiment, there will be some number of wins, so let's define a random variable Y_i that for a given trial i consisting of 36 throws will yield the value $Y_i = y_i$, where y_i is the number of wins in the i-th trial.

Our theoretical assumption that the probability of a win on any individual throw of the die is 1/6, and that this value is the same for each throw of the die and independent of the outcome of other throws (in statistical parlance, the outcomes of individual throws are '**I**ndependent and **I**dentically **D**istributed,' or **IID**) yields information about both facets of our experiment via the *binomial distribution*, which is a mathematical function of the number of opportunities (here $N = 36$) and the probability of winning in a single opportunity (here $p = 1/6$). We will forego the mathematical details here (this is covered in any elementary statistics or probability text), and simply display the results for the possible numbers of wins of Game 1 in a trial consisting of 36 throws of the die. The distribution is given in Figure 9.1. This distribution reveals that the mean value or *expectation value* for Y_i, $\langle Y_i \rangle_T = 6.0$ for each i (the subscript T indicates that this is the theoretical value). In other words, for each trial, the expected number of wins is 6. The distribution also yields a useful measure of variability, or the degree to which the results are 'spread out' around the expectation value. This is given by the *standard deviation* σ, which for the binomial distribution is given by the expression $\sigma = \sqrt{N}p(1-p)$ (≈ 2.24 for Y_i).

For comparison, Figure 9.2 shows the data from the 100 experiments by tabulating how many of them yielded particular values for Y. The observed mean of these experiments is $\langle Y \rangle_0 = 6.13$.

Our example of the dice game exhibits simply and obviously a feature that is found throughout the sciences in more complex and subtle ways: although what happens in any *individual* case is unpredictable, the data *in aggregate* exhibit an understandable pattern that can serve as a basis for forming expectations about the future. Many advances in statistics have

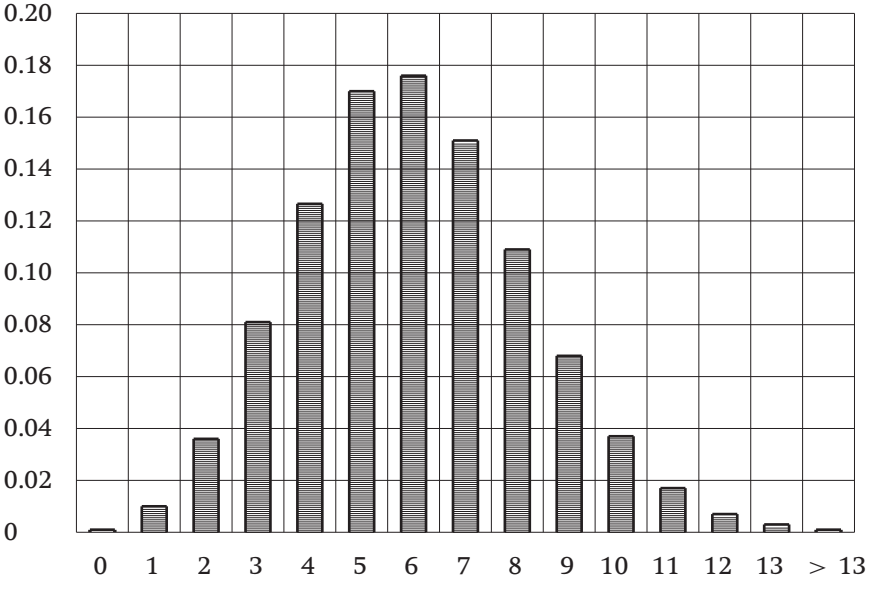


Figure 9.1 The binomial distribution for Y_i ($p = 1/6$, $N = 36$).

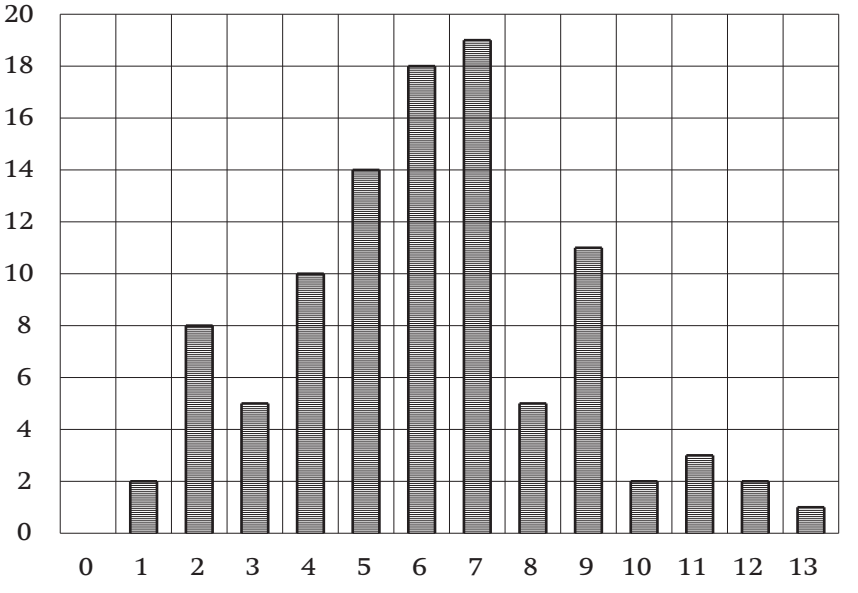


Figure 9.2 Dice game results, $N = 100$

come through the efforts of mathematicians and natural philosophers to duplicate the successes enjoyed dealing with games of chance in the analysis of astronomical data, in the prediction of the behavior of markets and other economic phenomena, and in many other domains in which patterns emerge at larger scales from noisy behavior at some smaller scale. Returning to Reichenbach's two functions of frequentism, probability ideas and the statistical science based upon them are so useful because they help both to *represent* such patterns, and to facilitate their *discovery*. Thus far we have emphasized the representation function; next we consider how concepts of frequentist statistics facilitate the discovery of such patterns.

9.3 Two traditions in frequentist statistics

In our die-throwing example, we have assumed that we know from the outset that the probability of winning Game 1 on a single throw of the die is 1/6, because we have assumed the die is fair. If, however, we did not know that, we would face quite a different situation. If I *know* that the die is fair, then it is reasonable to respond to my failure to win any games in the first 12 throws, and winning only one game in the first 36, by saying "what bad luck!" If, however, I do not already know that the die really is fair, then I might be inclined to respond to those same data by thinking that the data indicate not that my luck is bad, but that the die is not fair. But would such data constitute good evidence that the die is loaded?

In this section, we will consider two important schools of thought within the frequentist tradition in statistics, to see how they might address such questions. One tradition emphasizes the search for *statistically significant* discrepancies from some target hypothesis. The second tradition advocates the use of statistics to test among a class of hypotheses all belonging to some more general model. We will note the genuine differences between these approaches, but also consider a philosophical approach to scientific reasoning that draws on insights borrowed from both camps.

9.3.1 Fisherian significance testing

The kind of problem that interests us confronts not only those who like to gamble with dice, but also their kindred spirits, research scientists. Let's consider a recent example.

Example: the Higgs boson?

In July 2012, scientists at the Large Hadron Collider (LHC), a particle accelerator at the CERN (Conseil Européen pour la Recherche Nucléaire) laboratory near Geneva, Switzerland, announced they had discovered a new particle very similar, and perhaps identical, to the previously hypothetical Higgs boson (additional data allowed them later to conclude unambiguously that it was the Higgs). A discussion of what exactly the Higgs boson is and why it is such an exciting discovery would take us far afield of our immediate concerns (Weinberg, 2012). (Suffice it to say that in the prevailing Standard Model of the elementary constituents of the physical world, the Higgs boson is implicated by an ingenious hypothesis that seeks to explain why particles like electrons and quarks are not massless.) If the Higgs boson exists, then it will be produced out of the energy released in collisions between super-high-velocity protons, and then decay into other particles according to characteristic patterns dictated by well-established principles of the Standard Model. Two groups of physicists at the LHC use massively complex particle detectors to measure the properties of the particles produced in proton–proton collisions in order to see how well they match the 'signature' pattern expected from Higgs boson decays.

Two features of this scientific endeavor make statistics relevant. First, the theoretical description of the production of Higgs bosons rests in Quantum Field Theory, an inherently statistical theory yielding predictions in terms of probability distributions rather than certain outcomes. Second, *background* processes that will occur even if the Higgs boson does not exist at all can mimic the signature of Higgs decays, and these processes are also best described by probability distributions.

The two teams of LHC physicists, in announcing their discoveries, both described it as a "5-sigma (σ)" effect (Chatrchyan et al., 2012; Aad et al., 2012). Here σ refers to our just-introduced statistical concept the *standard deviation*.

Here is how that is relevant: They knew that even in the absence of the Higgs boson, they would find some number of Higgs-like 'signatures' in their data, a number they could only predict with probability. Part of their problem (a very difficult part) was to understand this *background distribution*, so that they could say, for any given number of Higgs-like signatures found in their data, what the probability is that they would get as many or more of such signature decays if only background processes were present. Their background

Table 9.2 *Shopping list for a Fisherian test.*

a question	based on what you want to learn about
a null hypothesis	a possible and testable answer to that question
data	from observations or experiment
a test statistic	some function of the data
	known probability distribution if the null hypothesis is true
	larger values indicate stronger evidence against null
the null distribution	the distribution of the test statistic assuming the null is true

distribution, just like the binomial distribution we used to characterize the behavior of a fair die, had an expectation value – the average number of signature events that would be expected from background alone – and a standard deviation – a measure of the variability of the background in terms of the *width* of the distribution, describing how much the number of signature events could be expected to vary from one data set to another. By describing their result as a "5σ" effect, the LHC physicists were indicating how far out in the background distribution their result is: it is five standard deviations greater than the expectation value assuming that only background processes are present.

In this way, they followed the methodology advocated by Ronald Fisher (1890–1962), who made significant contributions to evolutionary biology and genetics, as well as statistics. The essence of Fisher's methodology can be described in fairly simple terms (see Cox & Hinkley, 1974, ch. 3 for a careful introduction), but it requires certain things. Think of Table 9.2 as a shopping list for a Fisherian test. One begins with a *question*. To learn something that will help answer that question, you need to be able to state a substantive hypothesis that is a possible answer to that question, which you can test. This is the *null hypothesis*. The hard part comes next: figuring out how to generate *data* for such a test. You will need to be able to define some quantity that is a function of the data, called a *test statistic*, which will have a known probability distribution if that hypothesis is true, such that larger values of the test statistic should indicate stronger evidence of departure from what is expected if the null hypothesis is true. The probability distribution of the test statistic under the null hypothesis is the *null distribution*. The null distribution serves as a mathematical *model* of the null hypothesis, and what we test directly is the *statistical hypothesis* that the data are generated by a process characterized

by the null distribution. We test that claim by looking at the value of the test statistic: how probable is it that one would get a value as great or greater than the observed value, assuming the null hypothesis is true?

For example: if we wish to test whether a die is fair, then our statistical hypothesis will state that the data are drawn from a binomial distribution with $p = 1/6$ and $N = 36$. We could use the average number of 'wins' in our previously described 100 experiments by defining our test statistic to be $T \equiv |\langle Y_i \rangle - \langle Y \rangle_T|$, the observed value of which we denote $t \equiv |\langle y_i \rangle - \langle Y \rangle_T|$. The search for the Higgs boson required much more complicated data used to test the hypothesis that those data are drawn from the probability distribution that characterizes 'background physics.' The relevant random variable in that case is the difference between the number of signature events observed and the expectation value of the background distribution (the test statistic used has a slightly more complicated definition). In reporting the results of the Higgs search, the shorthand "5σ" stands for a certain probability statement: that were the null hypothesis true, the probability of getting as large an excess of signature events as they did (or more) is about 3×10^{-7}. This kind of probability statement, based on the observed value of the test statistic, is called a *p-value*, or sometimes the *observed statistical signficance* of the result.[4]

To appreciate the role *p*-values play in reporting an experimental result, compare the values cited in the Higgs announcements to what we should report were we to use the data from die-throwing reported above to test whether our die is fair. Recall that the observed mean $\langle y \rangle_0 = 6.13$, whereas theoretically we should expect a fair die to yield an average of 6 wins out of every 36 throws. Does the difference of $t = 0.13$ mean anything? The *p*-value of a difference of 0.13 is approximately 0.61.[5] That means that if you do an experiment of 100 trials with 36 tosses each, and you *are* using a fair die, you will get values for the random variable T as large as 0.13 or more *quite often* – on average, about 61 out of 100 times. The LHC physicists cite their very

[4] In the die-tossing experiment we are testing for divergence from the null hypothesis due to differences in either direction; a significant excess *or* a deficiency relative to the expectation value might lead us to reject the null hypothesis, and this is reflected in the calculation of the *p*-value. This is called a *two-sided* test. In the case of the Higgs boson, only excesses are considered relevant to the calculation of the *p*-value, because the test is designed to test for departures from the null with respect to the appearance of non-background processes. This is a *one-sided* test.

[5] This is based on what statisticians call a two-sided *t*-test.

small *p*-values to communicate that their data strongly indicate the negative claim that the null hypothesis of background-only processes does *not* yield a distribution that does a good job of describing their data. (That the data *do* indicate the existence of a Higgs boson rests on further features of the data beyond what is captured by the *p*-value.) Our rather *large* *p*-value for the die-throwing experiment, by contrast, clearly does not indicate that the null distribution is incompatible with the data. The result we obtained seems to be pretty typical of a fair die.

Does that mean that these data are good evidence that the null hypothesis is true? Fisher declared that "the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis" (Fisher, 1949, 16). Note that the null hypothesis makes a very exact statement: The probability of winning Game 1 in a single toss of the die is *exactly* 1/6. Hypotheses that attribute exact values to quantities are often called *point hypotheses*. The slight asymmetry among the six faces of the die (each does have a different number of dots) provides some reason to suspect that, construed as an exact statement, the null hypothesis is false, but whatever discrepancy from the exact truth of the null hypothesis might obtain is likely to be too slight to be of any practical interest to us, and is in any case not of a sort that can be revealed by the experiment we performed.

Although of widespread use among scientists, *p* values are subject to various misunderstandings and misuses.[6] A common misunderstanding is to think of the *p*-value as the probability that the null hypothesis is false. Frequentists deny that probabilities can be applied to hypotheses (unless the hypothesis can appropriately be regarded as the output of some chance mechanism). Hypotheses such as 'This die is fair' or 'Humans first used pottery for cooking approximately 20,000 years ago' are either true or false. No sense can be made of the relative frequency with which they are true. (Bayesians, of course, are perfectly happy to assign probabilities to hypotheses such as these, but would not endorse the *p*-values delivered by Fisherian significance tests as the correct means to calculate them.) Rather than the probability that the null hypothesis is false, a *p*-value reports the probability that one would

[6] See (D. E. Morrison & Henkel, 1970) for a historically significant collection of perspectives and (Ziliak & McCloskey, 2008) for a recent jeremiad.

observe a value for the random variable being considered that fits the null hypothesis as poorly or worse than the value actually observed, *assuming that the null is true*.

It is also sometimes alleged that in significance testing the null hypothesis is 'presumed true' or 'given the benefit of the doubt.'[7] But being chosen as a target for a significance test does not in itself indicate any judgment regarding the truth of a hypothesis. (We already have noted that a null hypothesis might be useful to test in spite of our having good reason to suspect that it is false, at least strictly speaking.) Point hypotheses yield well-defined probability distributions (*sampling distributions*) for the random variable of interest, and thus the means for calculating *p*-values. Typically, the aims of the test, i.e., what the investigator seeks to find out from the data, dictate the choice of null hypothesis. If you are interested in learning whether the die is loaded, it is more useful to test the hypothesis that it is not than to test a hypothesis that says that it is loaded to some specific degree. If you are interested in finding out whether the Higgs boson exists, it is more useful to test the hypothesis that it does not than a hypothesis that attributes particular characteristics to it (though testing the latter kind of hypothesis might also prove important to understanding the import of your data).

Finally, one should bear in mind that a result that is *statistically* significant may be *scientifically* or *practically* insignificant. One way in which this may happen comes from the fact that as the sample size increases, the sampling distribution tends to become more sharply peaked around the expectation value, so that smaller discrepancies become more improbable assuming the null hypothesis is true. If we were, for example, to obtain $t = 0.13$ from our die-rolling data on the basis of 10,000 experiments rather than 100, we would find that our result has a statistical significance of about 3×10^{-7} (just like the data from the search for the Higgs boson!). At that point we might justifiably claim to have evidence that the die is not *exactly* fair, but we might not care about such a small departure from fair odds. Indeed, with enough data, any slight discrepancy from the null expectation can become very statistically significant. With $N = 1,685,000$, a $p = 3 \times 10^{-7}$ result would be obtained with $t = 0.01$!

[7] Even well-known and mathematically sophisticated social scientists can fall into this error (Silver, 2012).

None of this means that p-values are valueless. The statistician David Cox summarizes the point nicely:

> The essential point is that significance tests in the first place address the question of whether the data are reasonably consistent with a null hypothesis in the respect tested. This is in many contexts an interesting but limited question.

Cox goes on to note that a more developed methodology will allow us to formulate "more informative summaries of what the data plus model assumptions imply" (Cox, 2006, 42). What might such a more developed methodology look like?

9.3.2 Neyman–Pearson testing

Fisherian hypothesis testing aims to discriminate between the null hypothesis and *every other* possible probability distribution. If a Fisherian test provides evidence against the null, that entails only that it provides reason to believe the data incompatible (statistically) with the null and compatible with *some other* hypothesis, without the latter being further specified.

Model assumptions

Consider again our Fisherian test of the hypothesis that the die is fair. The null hypothesis is well defined, but if we found the data statistically incompatible with the null hypothesis, we might want to know what we can reliably conclude from the discrepancy that we do find. For example, intuitively you might think that if our 100 experiments of 36 trials each yielded a value of $\langle Y \rangle_0 = 18.0$, then we could infer that $p \approx 1/2$, and that, more generally, we can infer, for whatever value we obtain for $\langle Y \rangle_0$, that $p \approx \frac{\langle Y \rangle_0}{36}$. This intuitive judgment is based, however, on an easily overlooked assumption that needs to be made explicit.

The quantity p is a parameter in the binomial distribution that we used (by letting $p = 1/6$) as a mathematical model of the hypothesis 'the die is fair.' That model, along with information about our testing procedure, was used to derive the sampling distribution for the random variable Y, *under the assumptions of the null hypothesis*. But if the data indicate that we have reason to *reject* the null hypothesis, what happens to those assumptions? In particular,

may we assume that the behavior of the die is best described by the binomial distribution? If not, then we have no business drawing conclusions about the value of p, which has meaning only as a parameter in that distribution.

As already noted, the binomial distribution assumes that the outcomes of individual throws are '**I**ndependent and **I**dentically **D**istributed,' or **IID**. The hypothesis that the die is fair involves not only attributing a probability to winning the game on a single throw of the die, but also asserts that that probability for a given throw is independent of the outcome of previous throws and remains the same for each throw in the sequence (the possible outcomes of each throw are described by the same – identical – probability distribution). So, the null hypothesis that picks out the value $p = 1/6$ belongs to a family of hypotheses. We can use the notation $Y \sim B(n, p)$ to mean 'random variable Y is distributed according to the *B*inomial distribution with parameters n and p.' Then we can express the statistical null hypothesis as $H_0 : Y \sim B(36, 1/6)$.

The intuition that we can infer, at least approximately, the value of p from the observed value $\langle Y \rangle_0$ really rests on the intuition that most dice satisfy the **IID** assumptions, even if we cannot always trust them to be fair. (Whether this intuition is warranted is, of course, another matter.) If we assume that there is some value of p for which $Y \sim B(36, p)$ is true for some value of p (we control the value of n, so we just *make* $n = 36$ true when we perform the experiment), then we can use the results of this test to draw conclusions about the range of values for p that are *supported* by those data. In other words, we assume the model $Y \sim B(36, p)$, and use that model to guide our inferences about the value of p. But in doing so, all of our inferences are hostage to the adequacy of the assumed model.

The question of the *adequacy* of the assumed model in model-based inferences raises important issues about how such models are specified and tested (Tukey, 1960; Mayo & Spanos, 2004; Spanos, 1999; Staley, 2012). Here we will focus on the question of how to use that model so as to ensure that we are drawing reliable inferences from the data in an optimal manner.

9.3.3 Error probabilities

In the late 1920s and early 1930s, Jerzy Neyman (1894–1981) and Egon S. Pearson (1895–1980) laid out a mathematical framework for choosing the *optimal* approach to testing a given hypothesis (Neyman & Pearson, 1928, 1933). Their framework, by invoking the kind of model assumptions just

discussed, allows for more definite statements regarding the comparison between the null hypothesis and its alternatives than a Fisherian test would support.

Again, we will need to forego a full mathematical discussion of Neyman–Pearson (NP) theory and settle for a characterization that will allow us to grasp the basic stance that this approach takes toward the problem of statistical inference (Cox & Hinkley, 1974, ch. 4 again provide a helpful systematic treatment).

Note that model assumptions such as **IID** delimit a class of possibilities. If our test procedure assumes at the outset that the die-tossing experiment outcomes are **IID**, then we will not be able to use that procedure to make inferences about hypotheses incompatible with those assumptions, such as a hypothesis according to which winning the game on one toss of the die makes it 50 percent less probable that you will win on the next toss. In the NP approach, model assumptions determine the class of hypotheses amongst which the test can discriminate. If the statistical null hypothesis is $H_0 : Y \sim B(36, 1/6)$, then the *alternative* against which we are testing it is $H_1 : Y \sim B(36, p); p \neq 1/6$. In other words, all hypotheses under consideration in this test assume that Y is binomially distributed with $n = 36$, but the null specifies one value for p, and the alternative says that p has some other value. As is typical in an NP test, the null hypothesis here is a point hypothesis ascribing a precise value to a parameter in the statistical distribution, whereas the alternative *compound hypothesis* simply says that the value of that parameter lies *somewhere* within a range of numbers – in this case the interval $[0, 1]$, exclusive of the point $1/6$.

Size, or significance level

Traditionally, NP tests are described as having two possible outcomes: accept the null H_0 or reject it. As Neyman and Pearson noted, having two possible outcomes means the test is susceptible to two distinct kinds of error:

> If we reject H_0, we may reject it when it is true; if we accept H_0, we may be accepting it when it is false, that is to say when really some alternative [H_1] is true. These two sources of error can rarely be eliminated completely; in some cases it will be more important to avoid the first, in others the second. (Neyman & Pearson, 1933, 296)

To specify an NP test for discriminating between H_0 and H_1, one begins by choosing the level at which one would like to limit the probability of the first kind of error, called a *Type I* error. For example, we may decide to design our test so that it will reject H_0 when it is true not more than once in 20 times. Then we may determine the *critical value* t_c for the test statistic T such that the probability of getting a value for T that is greater than or equal to t_c, assuming that the null hypothesis is true, is .05, i.e.,[8]

$$Pr(T \geq t_c; H_0) = .05. \tag{9.2}$$

More generally, one specifies an NP test to have a type I error probability that is limited to a certain value α (called the *size* or *significance level* of the test) by choosing a *critical value* t_c for the test statistic so that the condition

$$Pr(T \geq t_c; H_0) = \alpha \tag{9.3}$$

is met.

Power

The choice of a critical value t_c, then, ensures a certain kind of reliability for the test. By choosing a large enough value to make α, the Type I error probability of the test, very small, the investigator ensures that her test will only very rarely reject the null when it is true. One could, of course, ensure the same with a Fisherian test. The LHC physicists did just this in searching for the Higgs boson. They set a standard such that they only announce a discovery of a new phenomenon when the probability of such a large divergence from the null is $\sim 3 \times 10^{-7}$ or less, assuming that the null hypothesis is true.

[8] The appearance of a semicolon in Equation (9.2) introduces some new notation. Expressions of the form '$Pr(A;B)$' should be read as 'the probability of A assuming that B is true.' This verbal expression might not seem substantively different from the conditional probability statement 'the probability of A given (or conditional on) B,' which we notate '$Pr(A|B)$'. However, mathematically speaking, the difference is significant. Recall that conditional probabilities are defined such that $Pr(A|B) = Pr(A \wedge B)/Pr(B)$. (This is the version of Definition 2 (Conditional probability), from Chapter 8, for probabilities applied to sentences. The symbol '$\wedge$' represents logical *conjunction*, which connects two sentences into a single sentence that is true if and only if both of the conjoined sentences are true.) For frequentists however, the conditional probability expression $Pr(T \geq t_c|H_0) = Pr(T \geq t_c \wedge H_0)/Pr(H_0)$ would be meaningless. There is no relative frequency with which H_0 is true, making both the numerator $Pr(T \geq t_c \wedge H_0)$ and the denominator $Pr(H_0)$ undefined.

But what about the other kind of error, in which one accepts the null hypothesis although the alternative is true? This is called *Type II* error. One of Neyman and Pearson's innovations was to introduce a quantity called the *power* of a test, which is one minus the probability of a Type II error. This enabled them to address what they saw as a shortcoming in Fisher's approach, namely, that Fisher did not provide a systematic rationale for the choice of test statistic.

Neyman and Pearson tackled, in a statistical framework, a general problem that inquirers confront. On the one hand, we do not want to acquire a large stock of false beliefs. We would like to reason *reliably*. On the other hand, we also do not want to fail to learn things that are true. One could always avoid the acquisition of false beliefs by simply not forming beliefs. But having *true* beliefs is valuable to us for many reasons.

So a test should not only have a small size or significance level, it should also have high power, which is to say, a high (enough) probability of detecting an effect of interest, if it is present.

Given that we assume at the outset of our test that we are discriminating between the null hypothesis H_0 and the alternative H_1, you might think that the power of a test can be defined as $1 - Pr(T < t_c; H_1)$. This would indeed be the case were one to test the null point hypothesis against a *point* alternative hypothesis, because then the probability distribution needed for determining the value of $Pr(T < t_c; H_1)$ would be specified. Rarely, however, do real scientific problems lend themselves to such a treatment. Much more typical are cases where the target null hypothesis is contrasted with a continuum of alternative possibilities, just as in our die-rolling example. In such cases, the power of the test varies according to the various possible alternatives in H_1.

What Neyman and Pearson's work allows investigators to do, then, is to define a test, both in the choice of the critical value of the test statistic and in the choice of the test statistic itself, that is optimal in a certain sense: For a given size (Type I error probability), it is in many cases possible to choose a test that also maximizes the power (minimizes the Type II error probability) for *all* possible alternatives in H_1 (often called a *uniformly most powerful test*). Just how one goes about determining the optimal test is an important problem for the specification of NP tests, but would take us too far afield of our present concerns. We now turn instead to the questions of how to understand the aims of NP testing and how to interpret the outcomes of NP tests. We will consider some traditional interpretations and their shortcomings before turning

to Deborah Mayo's error-statistical philosophy that seeks to overcome those limitations through a reconsideration of both the aims and methods of frequentist statistical appraisal.

9.3.4 Statistical tests: what are they good for?

A behavioristic interpretation

In their 1933 paper, Neyman and Pearson articulate a rationale for NP testing that emphasizes the usefulness of such tests for limiting the rate of erroneous decisions in the long run. Having asserted that "no test based upon the theory of probability can by itself provide any valuable evidence of the truth or falsehood" of a particular hypothesis, they go on to note:

> But we may look at the purpose of tests from another view-point. Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong. ... Such a rule tells us nothing as to whether in a particular case H is true when [the rule says to accept it] or false when [the rule says to reject it]. But it may often be proved that if we behave according to such a rule, then in the long run we shall reject H when it is true not more, say, than once in a hundred times, and in addition we may have evidence that we shall reject H sufficiently often when it is false. (Neyman & Pearson, 1933, 290–291)[9]

Here Neyman and Pearson express a *behavioristic* understanding of the rationale for NP testing, according to which the results of an NP test do not indicate whether data provide good support or evidence for or against a given hypothesis. Rather, on this interpretation, the NP tester begins with a choice between two courses of action that provide the practical signification of the outcomes *accept H* and *reject H*, as well as some idea of the costs of making the choice between those actions erroneously. Following the decision rule specified as part of the NP test, by allowing for the limitation of error rates, allows the NP tester to contain costs. Indeed, by facilitating the design of

[9] Neyman continued in later works to advocate an interpretation of NP tests in terms of "inductive behavior" (Neyman, 1950, 1–2), but Pearson, argues Deborah Mayo, became a "heretic" to the Neyman–Pearson philosophy and advocated a view closer to Mayo's own (Mayo, 1996, ch. 11).

optimal tests, Neyman and Pearson allow one to maximize a certain kind of *efficiency*.

Critics (such as R. A. Fisher) have noted that, whatever usefulness NP tests thus understood might have for, say, the pursuit of economic gains in manufacturing (a typical example would be the use of sampling from batches of manufactured goods to see whether they meet quality-assurance standards before shipping them to buyers), the relevance of such considerations to science is doubtful. Scientists do seek answers to just the kinds of questions that Neyman and Pearson say they cannot answer. They would like to have "valuable evidence of the truth or falsehood" of the hypotheses they investigate!

A measure-of-evidence interpretation

An alternative to such a behavioristic interpretation would be to interpret the error probabilities of a test as indicators of the strength of evidence provided by the test with regard to the hypotheses tested. An approach of this kind, which we might call a *measure-of-evidence* interpretation, was attempted by Allan Birnbaum (Birnbaum, 1977). In Birnbaum's approach, a rejection of the null hypothesis in favor of the alternative (or vice versa) would be interpreted in terms of the strength of evidence it provided ('strong,' 'very strong,' 'conclusive,' 'weak,' or 'worthless') in favor of the alternative as against the null (or vice versa). This interpretation would be based simply on the error probabilities of the test that led to that rejection. For example, if a test yielded a rejection of H_0 in favor of H_1 with error probabilities $\alpha = 0.06$ and $\beta = 0.08$ (H_1 must be a point hypothesis in order to yield a single value for β), then Birnbaum asserts he would interpret that as "strong statistical evidence" for H_1 as against H_0 (Birnbaum, 1977, 24).

As Birnbaum acknowledges, statistical evidence in the form of outcomes of statistical tests alone do not suffice to form reasoned judgments regarding substantive scientific questions. Even as an account of strictly statistical evidence, however, Birnbaum's approach is inadequate. Two different tests with the same accept/reject outcome and the same error probabilities need not yield statistical evidence of the same import. For example, in an NP test, one designates a single cutoff value t_c, and the test rejects the null hypothesis if the data yield any value for T that is greater than t_c. Neither the outcome nor the error probabilities α and β are sensitive to the magnitude by which the observed value of T exceeds the cutoff value. For example, setting a cutoff value $t_c = 0.5$ for testing the fair die hypothesis on the basis of the 100 trials

of 36 tosses yields a size of $\alpha \approx .05$. The null hypothesis H_0 would then be rejected by data yielding observed values of $t = 0.6$ as well as by data yielding $t = 3.6$, but the latter result gives much stronger evidence against the null than the former. Also, as mentioned previously, as the sample size grows, the sampling distribution tends to become more sharply peaked around the expectation value. As a consequence the null will be more easily rejected at a given size by a smaller effect if the sample size is large than if it is small. This is another way in which factors beyond the accept/reject outcome of a test and its error probabilities have a bearing on the interpretation of the statistical evidence (see Mayo, 1985, 1996; Pratt, 1977).

9.4 Error-statistical philosophy of science and severe testing

Deborah Mayo has developed an alternative approach to the interpretation of frequentist statistical inference (Mayo, 1996).[10] But the idea at the heart of Mayo's approach is one that can be stated without invoking probability at all. Let's start with this core idea, and then we can see how and why probabilities enter the discussion.

Mayo takes the following "minimal scientific principle for evidence" to be uncontroversial:

Principle 3 (Minimal principle for evidence) *Data* $\mathbf{x}_0$ *provide poor evidence for H if they result from a method or procedure that has little or no ability of finding flaws in H, even if H is false.* (Mayo & Spanos, 2009, 3)

This minimal principle is weak: it states neither a criterion the satisfaction of which suffices for data to constitute *good* evidence, nor a criterion the satisfaction of which is necessary for data to constitute *poor* evidence. However, it does commit one to regarding not just the data themselves but the "method or procedure" by which those data were generated as *relevant* to their evidential import. Moreover, it indicates that the relevant aspect of that method or procedure concerns the "ability" it has to reveal a flaw in H, if there is such a flaw (such as being false). Mayo thus considers even this weak principle philosophically and methodologically significant.

[10] Mayo's 1996 book *Error and the growth of experimental knowledge* is the most comprehensive of Mayo's publications advocating her approach, and gives her perspective on many issues discussed previously in this text. The introduction to a more recent collection edited by Mayo and her collaborator Aris Spanos includes an incisive summary of the error-statistical philosophy (Mayo & Spanos, 2009).

Philosophical accounts of scientific reasoning have, however, generally failed to satisfy this principle.[11] Philosophers of science have constructed theories of evidence that divorce consideration of the evidential import of data from consideration of the methods used to generate those data. Mayo argues that we should *reject* the project of the logical positivists that aimed to allow one, for any data or observations E, to calculate the degree of support or confirmation afforded to any hypothesis H. We should not, however, abandon the ideals of *neutrality* and *objectivity* themselves. An account of scientific reasoning that respects Principle 3 will better promote these ideals by emphasizing that reliable inferences from data require consideration of the properties of the method that produced them.

Principle 3 tells us that the properties of methods that matter to the evidential appraisal of data are those relevant for a method's "ability of finding flaws" in the hypothesis under consideration. Error probabilities enter into Mayo's account because she maintains that (1) probability distributions provide useful models of the operating characteristics of methods of hypothesis appraisal, and (2) the operating characteristics that best reflect a method's "ability of finding flaws" in relevant hypotheses are its error probabilities as calculated from such a probability distribution (assuming that distribution is in fact an adequate model of the method).

9.4.1 Severity and evidence

The use of probabilities as models for the flaw-finding capacities of tests leads to a refinement of Principle 3 that addresses the question: 'When do data $\mathbf{x}_0$ provide good evidence for or a good test of hypothesis H?'

Principle 4 (Severity principle) *Data $\mathbf{x}_0$ provide a good indication of or evidence for hypothesis H (just) to the extent that test T has severely passed H with $\mathbf{x}_0$. (Mayo & Spanos, 2009, 21)*

[11] An exception would be the account given by Charles S. Peirce (1839–1914), one of the founders of the philosophical approach *pragmatism*. Peirce insisted that in *probable reasoning* one must take account of "the manner in which the premisses have been obtained." Peirce's theory of probable reasoning anticipates some features of the Neyman–Pearson approach and Mayo's error-statistical elaboration of it (Peirce, 1883, 128; see also Kyburg, 1993; Mayo, 1993).

We need to understand what the expression "severely passed" means, but first note that this expression relates *three items*: (1) the hypothesis under test, (2) the data with which the hypothesis passes, and (3) the test itself. Non-error-statistical accounts of evidence or confirmation typically only concern themselves with a relation between the data and the hypothesis.

Mayo gives the following account of what it means to pass a severe test or to pass a test with severity:

Definition 4 (Severity) *A hypothesis H passes a severe test T with data* $\mathbf{x}_0$ *if*

(S-1) $\mathbf{x}_0$ *agrees with H (for a suitable notion of "agreement") and*
(S-2) with very high probability, test T would have produced a result that accords less well with H than does $\mathbf{x}_0$*, if H were false or incorrect. (Mayo & Spanos, 2009, 22)*

Evaluating the "agreement" in condition S-1 will often use probabilities: the data fit or agree with the hypothesis H well if H confers a probability on the observed value of the test statistic (or whatever data-dependent quantity serves as a measure of distance from what is expected if H is true) that is relatively high compared to the probabilities conferred by competing hypotheses.

As a *first pass* at appreciating the relevance of condition S-2, consider the context where the hypothesis that is considered to pass the test with data $\mathbf{x}_0$ is the alternative H_1 in an NP test of the fair die hypothesis. If (1) our test has a cutoff fixed at t_c that results in a test with significance level α (= .05, let's say), (2) the data yield a value of $t = t_c$, and (3) we take this to be a passing result for the alternative hypothesis H_1, then perhaps we can use the significance level of the test to evaluate how well condition S-2 has been met. Specifically, we know that $Pr(T \geq t_c; H_0) = \alpha = .05$. In other words, assuming the null hypothesis is true (and hence the alternative false), it would rarely (once in twenty times, on average) happen that one would get a result that agrees so poorly (or even worse) with H_0. It follows then that $Pr(T < t_c; H_0) = 1 - \alpha = .95$. In other words, assuming the null hypothesis true and the alternative false, one would almost always obtain a value for T that agrees better with the denial of H_1 (i.e., with H_0) than does the result in hand. So can we say that conditions S-1 and S-2 have been met by H_1 and it has passed a severe test with such a result?

Answering requires us to go beyond the standard NP testing approach we have invoked so far in this example. Recall that H_1 is usually (as in the die example) a *compound* hypothesis. In our example it is $H_1 : Y \sim B(36, p); p \neq 1/6$. Although this hypothesis says something definite about the kind of

distribution that describes Y, all that it says about the value of the parameter p is $p \neq 1/6$, which is not saying very much at all! This makes the idea of "agreement" with H_1 so poorly defined as to preclude the application of the severity concept. If we want to know what hypotheses might have passed with severity in our example, we will have to scrutinize the various possibilities buried within the compound hypothesis H_1.

Recall that the original rationale for NP tests appeared to be a behavioristic one focused on limiting the rate of erroneous decisions, whereas we now seek an account of what we can learn about hypotheses from experimental data. Mayo insists that the NP emphasis on the size and power of a test remains relevant, particularly for *pre-data planning*. Such considerations help investigators devise the most efficient ways to gather data to facilitate learning about the hypotheses under test. But given that size and power do not translate directly into a measure of the strength of evidence for or against a hypothesis, one must use a broader consideration of error probabilities to discriminate amongst the possible inferences one might draw with regard to the entire family of hypotheses under consideration *after the data have been gathered*.

Before turning to an example to see how we may implement this idea, two points deserve emphasis. *First*: remember that probabilities entered into the account as a means for representing the error-detecting capacities of test procedures. Very strong arguments based on severity considerations can often be given without appealing to quantitative probability measures at all. *Second*: Mayo proposes that we may regard learning from experimental data as a matter of learning to construct *arguments from error*, the structure of which "is guided by the following thesis":

Principle 5 (Arguing from error) *It is learned that an error is absent when (and only to the extent that) a procedure of inquiry (which may include several tests) having a high probability of detecting the error if (and only if) it exists nevertheless fails to do so, but instead produces results that accord well with the absence of the error.* (Mayo, 1996, 64)

So we can apply error-statistical reasoning without explicit appeals to error probabilities. Consider the example from Chapter 1, in which Newton investigated the differential refrangibility of sunlight. An argument from error would require us to ask first, what are the ways in which it might be erroneous to conclude – as Newton does – that light from the Sun consists of rays that are differentially refrangible? We would then want to ask whether

Newton's method of inquiry had a high probability (or a strong ability) of revealing such errors, supposing them to be present. In doing so, we can determine which conclusions are warranted by having been well-probed for possible errors (e.g., refraction with the prism reveals sunlight to be analyzable into light of different colors, with the different colors refracted at different angles; the refracted light is not further analyzable by the same means into further components), as well as which conclusions are not warranted because the method is incapable of revealing them as erroneous even if they are. (Such refraction does not warrant any conclusions about whether light is composed of particles or waves, for example. Newton's experiment could not reveal the flaws in such claims, should there be such flaws.)

9.4.2 Example: evidence that formaldehyde is carcinogenic

Let us now develop an example in more detail.

Formaldehyde is a colorless flammable gas found in a wide range of products including some cosmetics, cleaning products, and carpets, as well as in cigarette smoke and some industrial exhaust. If you have ever dissected a deceased frog or other creature preserved in formalin (a solution of formaldehyde gas in water), then you know its distinctive odor. The US Environmental Protection Agency (EPA) presently lists formaldehyde as a "probable human carcinogen, based on limited evidence in humans, and sufficient evidence in animals" (US Environmental Protection Agency, 2012). The National Toxicology Program of the US Department of Health and Human Services (DHHS), in the Twelfth Edition (2011) of its *Report on Carcinogens*, revised the status of formaldehyde from "reasonably anticipated to be a human carcinogen" (as it had been listed since 1981) to "known to be a human carcinogen," citing "sufficient evidence of carcinogenicity from studies in humans and supporting data on mechanisms of carcinogenesis" (US Department of Health and Human Services Public Health Service, 2011, 195).

Much of the evidence that formaldehyde causes cancer in humans comes from epidemiological studies, i.e., studies that compare rates of various types of cancer among humans known to have significant exposure to formaldehyde (pathologists and embalmers, for example) with rates among humans without such exposure. Such studies require great care to ensure that the selection of subjects and interpretation of data is unbiased, and to take into account possible confounding causes (other causes of cancer, exposure to

which may differ among the two groups studied). Moreover, measuring actual rates of exposure to formaldehyde is difficult, even among those whose jobs involve regular use of formaldehyde-containing substances. The recent determinations that formaldehyde is indeed a human carcinogen, as reflected in the DHHS *Report on Carcinogens*, result from significant advances in the ability of epidemiologists to tackle these problems, as well as the accumulation of data over many years, allowing the long-term effects of formaldehyde exposure finally to emerge.

In the early 1980s, the situation was rather different. The few existing epidemiological studies were generally considered inconclusive. Experiments on rats and mice, however, yielded more definite results. Our discussion of these experiments will focus on how scientists used frequentist statistics in the evaluation of their evidence, and how an error-statistical approach would interpret and possibly expand upon those assessments.

These experiments and the epidemiological studies that had been carried out at the time were central in a controversy that arose over the classification of formaldehyde in the early 1980s. Although several scientific panels and the EPA's own scientists had recommended that the substance be classified as a carcinogen, EPA administrators who were newly appointed by the administration of President Ronald Reagan declined to do so. Critics charged the EPA administrators with having neglected scientific evidence to advance a political agenda of de-regulation of industry (see the discussions in Ashford, Ryan, & Caldart, 1983; Jasanoff, 1987 and Mayo, 1991).

In experiments carried out by researchers working for the Chemical Industry Institute of Toxicology, male and female rats and mice were randomly assigned to four groups subjected to different levels of exposure to formaldehyde gas: 14.3, 5.6, 2.0, and 0 parts per million (ppm). Each exposure group had about 120 members of each of the two species and each of the two sexes. Those exposed to 0 ppm were the *control group*, with whom comparisons would be drawn. Exposures were scheduled to last for 24 months, followed by 6 months of nonexposure. At intervals of several months, researchers would 'sacrifice' pre-established numbers of animals; they would then conduct pathological examinations of these and other study animals that had deceased.

We will not concern ourselves with the subtleties of the statistical analysis of the data from these studies, but will content ourselves with a rough characterization of the hypotheses under investigation and the nature of the support they established for the inferences they drew. Although the researchers

report a number of pathological observations, we will focus on results that became central to debates over the safety of formaldehyde: the incidence of squamous cell carcinomas (SCCs) in the various groups.

First let us clarify what hypotheses the researchers tested with their data on the incidence of SCCs. Most directly, they tested hypotheses concerning the *difference* that exposure to formaldehyde makes to the risk of developing SCCs in rats and in mice. A null hypothesis regarding this difference would state that formaldehyde makes no difference to the risk of developing SCC. We could define a parameter RI_{R_E} to be the difference between the risk of SCC among rats exposed to formaldehyde at a level of exposure E and the risk among those not exposed to formaldehyde. Then the null hypothesis would be $RI_{R_E} = 0$.

Among rats in the 14.3 ppm exposure group, 103 were found to have SCCs, as compared to 2 in the 5.6 ppm group, and none in either the 2.0 ppm or 0 ppm group. Among mice, investigators found 2 SCCs in the 14.3 ppm exposure group, and none in the other groups. How should we interpret these results?

Following a widespread convention, the researchers do not report actual p-values for these but instead report the results as either surpassing or failing to surpass certain thresholds of statistical significance. They note that the 103 SCCs in the 14.3 ppm exposure rats are significant at a level of $p < 0.001$. As is common practice in the biomedical sciences, the researchers report results that have p-values above 0.05 as 'not significant.' The 2 SCCs in the 5.6 ppm exposure rats and the 2 in the 14.3 ppm exposure mice fall into this category.

The researchers use 'significant' as a technical term, but it is easily misunderstood. It does not mean the same thing as 'important,' either when it is affirmed or when it is denied of a given result. To say that finding 103 SCCs among the 14.3 ppm exposure group is *significant at a level of* $p < 0.001$ is to say that, assuming $RI_{\mathrm{R}_{14.3ppm}} = 0$, one would expect to find 103 or more SCCs in a similar sample of rats less than 1 in 1,000 times. (This does *not* mean that there is a less than 1 in 1,000 chance that these SCCs were due to chance or that there is a less than 1 in 1,000 chance that formaldehyde does not cause SCCs in rats exposed at a level of 14.3 ppm. Such statements would concern the probability of a particular explanation of the data, which is not licensed in a frequentist framework.) To say that the 2 SCCs in the 5.6 ppm exposure rats or the 2 SCCs in the 14.3 ppm mice are 'not significant' simply means that, assuming $RI_{R_{5.6ppm}} = 0$, finding 2 or more SCCs would happen more often than

once in 20 times, on average. These latter two results may not be significant in the technical sense, but as we will see they are *relevant*.

Regarding the much greater incidence of SCCs among rats exposed at 14.3 ppm as compared to that of mice exposed at the same level, the researchers note that "It is of interest to note that the incidence of squamous cell carcinoma was similar in mice exposed to 14.3 ppm and rats exposed to 5.6 ppm of formaldehyde" (Kerns et al., 1983, 4388). Although neither number is statistically significant, the similarity in the two rates *suggests* a hypothesis that would, if correct, explain the discrepancy between the effects of formaldehyde in the two species, namely, that the "differences in response … may be related to differences in their physiological responses to formaldehyde inhalation" that would result in rats receiving, at a given level of exposure, a much greater "dose" of formaldehyde than mice exposed at the same level (Kerns et al., 1983, 4388). The researchers cite independent support for such a hypothesis on the grounds of physiological differences in the nasal cavities of the two species as well as from data on "cell turnover in the nasal cavities of rats and mice" (Kerns et al., 1983, 4388).

From the standpoint of the Severity Principle, we might at this point wish to say that the data provide good evidence for the hypothesis that $RI_{R_{14.3ppm}} > 0$, since the difference in rates of SCCs among the control group and the 14.3 ppm exposure group is of a size that fits well with the hypothesis that formaldehyde exposure at that level increases the risk of SCCs, and it is highly improbable that one would observe as great a difference (or greater), were it the case that formaldehyde exposure did not have such an effect.

Metastatistical severity analysis of the fates of rats and mice

Mayo and Spanos (2006) have emphasized that an assessment of the severity with which a hypothesis has passed a given test can not be read off directly from *p*-values or the outcome of an NP test, but requires instead a *metastatistical severity analysis* of the results of such tests. Under the behavioristic rationale we discussed above, Fisherian or NP tests have their value in the ability the investigator has to limit her rate of erroneous inferences in repeated applications of the tests. But the rationale for a severity analysis is to determine what warranted inferences the investigator may draw from a *particular* set of data.

Mayo and Spanos specify three requirements for success at that task: (1) The analysis must concern inferences about the presence or absence of errors, rather than (merely) decisions about a course of action to pursue. (2) One must use a test statistic (such as rate of SSCs) that will serve as an "appropriate measure of accordance or distance" with respect to the hypotheses under consideration. Roughly, this is simply the idea that larger values of the quantity used to determine whether the test rejects or accepts a hypothesis H must indicate larger discrepancies from (or worse fit with) what one would expect were H true. (3) The assessment must be "sensitive to the *particular outcome* $\mathbf{x}_0$; it must be a *post-data assessment*" (Mayo & Spanos, 2006, 329–330).

In our example, a severity analysis would not, therefore, rest with simply reporting that finding 103 SCCs among the rats exposed at 14.3 ppm was significant with $p < 0.001$. One would next proceed to ask about hypotheses asserting specific sizes of increased risk, relying on the very same data. For example, do those data support the inference that $RI_{R_{14.3ppm}} > 0.10$ (i.e., that exposure at 14.3 ppm adds *more than 10%* to the probability that a rat will get an SCC)? The answer to this depends on how probable it is (how often it would happen) that one would obtain *no more than* 103 SCCs in this sample were it the case that exposure added *just* 10% to that risk. What about a 20% increase? 30%? and so on … In this way, one could calculate the severity, for a given possible inference about $RI_{R_{14.3ppm}}$, with which that hypothesis passes, given the data in hand. Although Mayo and Spanos do not propose any particular fixed severity threshold above which one should take the data to provide good support for a hypothesis and below which one should not, the idea is that, assuming that the above conditions for a severity analysis have been met, one can distinguish well-warranted from poorly warranted inferences based on whether the inferred hypotheses passed with high or low severity, respectively (Mayo & Spanos, 2006, 341–344).

Such an analysis can be especially useful, according to advocates, in the case that a test fails to reject the null hypothesis, or a result is found not to be statistically significant. What, for example, follows from the fact that among the mice exposed at 14.3 ppm, only 2 SCCs were found, which was judged to be not statistically significant? If we consider this as a passing result for the hypothesis $RI_{M_{14.3ppm}} = 0$ (in effect, 'formaldehyde is safe for mice'), does this constitute good evidence for that hypothesis? A severity analysis would

direct our attention beyond simply the absence of a statistically significant departure from what we would expect were this null hypothesis true. We would first ask how probable it is that the difference between the expected value under the null and the observed value of the test statistic would have been *larger* than it is, if $RI_{M_{14.3ppm}} > 0$. If that probability is *very high*, then the null hypothesis passes with high severity. But notice that here again, the calculation of this probability will depend on *which* value of $RI_{M_{14.3ppm}}$ we choose that is greater than 0. Presumably, if $RI_{M_{14.3ppm}}$ exceeds 0 by only a very small amount, such as 0.001, then it is *not* the case that it is highly probable that more than 2 SCCs would have been observed (Mayo & Spanos, 2006, 337–339).

In the case of a test that fails to reject a null of the form $\mu \leq \mu_0$ (or $\mu = \mu_0$), for some numerical value μ_0, a severity analysis therefore focuses on the severity of inferences of the form $\mu \leq \mu'$ for various values of μ' greater than μ_0. In our example, the question to ask of the mice results is what inferences they warrant of the form 'the increased risk in mice of SCCs from exposure to formaldehyde at 14.3 ppm is *not greater than r*.' Error statisticians thus acknowledge Fisher's cautions against accepting the null hypothesis based on a failure to reject it, but they go beyond Fisher insofar as they allow for a post-data analysis of hypotheses about the extent to which the data support inferences that the null is not wrong by more than a given amount. (This goes beyond orthodox NP testing as well, though Mayo and Spanos have argued that aspects of this approach can be found in the writings of Neyman himself (Mayo & Spanos, 2006, 334–336).)

A severity analysis requires one to distinguish between those errors that the results of a given test do and do not rule out (in the sense that, were those errors actual, the test would have the capacity to tell us so). Such a perspective would be potentially relevant to the broader significance of these results. What do they suggest about the health effects of formaldehyde on human beings?

One of the newly appointed EPA administrators involved in the early 1980s decision against regulating formaldehyde as a carcinogen noted that, although"formaldehyde is a carcinogen in the rat by the inhalation route … its carcinogenic potential appears to vary significantly with species and route." He furthermore stated that epidemiological data supported "the notion that any human problems with formaldehyde carcinogenicity may be of low incidence or undetectable" (Ashford et al., 1983, 327). We have

already seen that the investigators who performed the experiment on rats and mice suspected that mice might suffer lower rates of SCCs because of a difference in the dose of formaldehyde that they actually received, rather than (or perhaps in addition to) a difference in the carcinogenic potential of formaldehyde between mice and rats. In other words, the difference in the results between the two species could *not* be reliably interpreted as supporting a substantive claim about the difference in carcinogenicity in the two species, a claim which would not have been well tested by this experiment. Moreover, from the standpoint of a severity analysis, the absence of statistically significant results from epidemiological studies would only warrant the claim that the carcinogenicity of formaldehyde in humans is 'of low incidence or undetectable' were it the case that those studies had very high probability of finding an effect if the carcinogenicity of formaldehyde in humans exceeded the threshold of 'low incidence.' There was at the time good reason to think that this criterion was not met by these studies (Mayo, 1991, 271–273). In other words, even were it the case that formaldehyde posed a significant carcinogenic potential to humans, it was not improbable that these studies would have failed to find a significant effect.

9.5 Conclusion

In this chapter, we have considered the idea of relative frequency as a way of understanding probability statements, as the foundation for some important approaches to the statistical analysis of data, and as a jumping-off point for an approach to philosophy of science that regards the error-probing power of tests as crucial to the determination of what it is that we learn from scientific inquiry.

The error-statistical approach that we have explored here has been subjected to considerable criticism from other philosophers of science (Achinstein, 2010; Howson, 1997; Musgrave, 2010), and Mayo has responded to these criticisms in turn (Mayo, 1997b; Mayo & Spanos, 2009). Here I discuss two critical issues.

Some critics argue that the error-statistical approach is committed to regarding some hypotheses as passing severe tests which nonetheless have low posterior probabilities (Achinstein, 2010). Suppose that we are interested in the hypothesis H_T: Tatyana is well-qualified to participate in competitive eating contests. The contrary hypothesis H'_T states that Tatyana is not thus qualified. To test H_T against H'_T, we ask Tatyana to eat 10 saltine crackers in

60 seconds without drinking any liquids. If she can do so, she passes the test. Only one out of twenty people who are not well-qualified for competitive eating can pass this test, whereas nearly everyone who is thus qualified would pass it. However, in the city where Tatyana lives, only one person in a thousand is well-qualified for competitive eating. From the test we obtain the result E: Tatyana passes the test. We have the following probabilities:

$$\begin{array}{rclrcl} Pr(H_T) & = & 0.001 & Pr(H'_T) & = & 0.999 \\ Pr(E|H_T) & \simeq & 1 & Pr(E|H'_T) & = & 0.05 \end{array} \tag{9.4}$$

Suppose that we regard H_T as having passed a (somewhat) severe test. The severity principle would then have us regard this as (somewhat) good evidence for H_T. But a Bayesian calculation reveals that $Pr(H_T|E) \simeq 0.02$, which is still quite low, while $Pr(H'_T|E) \simeq 0/98$.

Mayo's response points out first that this criticism begs the question by invoking a criterion of appraisal rejected by error-statistics. That approach does not appraise a scientific claim by considering its posterior probability, but by asking how well it has survived scrutiny by methods with a good capacity for revealing errors, should they be present. Indeed, a test based on the criterion of posterior probability would use the results of Tatyana's performance as an indication that she is not well-qualified for competitive eating regardless of what she did, simply on the basis of the city where she lives, and would thus be committed to "innocence by association" (Mayo & Spanos, 2009, 192–201). (Spanos, 2010 offers another criticism of such arguments.)

The error-statistical approach does not merely offer a different answer to a traditional question in philosophy of science: 'how do we evaluate the degree to which a given body of observations confirm (make probable) or disconfirm (make improbable) a theory?' Rather, the error-statistician seeks to *replace* this question with another: 'how do we learn about scientific claims by subjecting them to tests designed to reveal such errors in those claims as we might be interested to discover?'

A second challenge for error-statistics has been to show how it can shed light on scientific claims that appear to be less directly connected with particular bodies of experimental data, what philosophers have sometimes called 'high-level theories.' Granting that a severity analysis provides a helpful perspective on the use of data from rats to evaluate the carcinogenicity of formaldehyde, what can it tell us about the status of such broad theoretical frameworks as Neo-Darwinian Evolutionary Theory, Keynesian Economic

Theory, or the General Theory of Relativity? These theories do not lend themselves to severe testing directly.

Although much work on this challenge remains to be done, the basic approach that defenders of error-statistics have taken is to argue (analogously to their response to the first challenge) that an understanding of how severe tests are relevant for high-level theories requires that one first *give up* the assumption that the nature of the problem concerns *how probable* such theories are or the assignment of a *degree of confirmation*. Instead, we should ask of our theories *the extent to which we have effectively probed them for errors and how well they have survived such probing*.

To date, Mayo's writings on this topic have focused on the case of experimental testing of theories of gravity (Mayo, 2002; Mayo & Spanos, 2009; see Staley, 2008, 2013 for a discussion and extension). In particular she has focused on the use of a framework in which Einstein's General Theory of Relativity (GTR) can be put into a common mathematical representation with other theories of gravity and tested, in a piecemeal way, for how well they describe a broad range of phenomena in the domain where relative velocities are not large and gravitational fields are not extraordinarily strong. In this way, data from many different kind of observations and experiments can be used to discriminate the various ways in which GTR *might* be in error (and other theories more correct), so that, although GTR might not be subjected in its entirety to a single severe test, we can severely test claims that it makes about, for example, how much a given quantity of mass curves the space around it, one of the features of GTR that distinguishes it from some of its alternatives (Will, 1993a, 1993b).

A significant feature of this case is that the theories to which this framework compares GTR are neither all of theories that are logically possible (which would include many that are silly or incomprehensible), nor only those theories that we are clever enough to have thought of (which might not include the correct theory). The framework allows one to test among an entire class of theories satisfying certain weak assumptions.[12] This includes

[12] Weak, but not trivial. The framework, called the Parametrized Post-Newtonian (PPN) framework includes all theories that satisfy the Einstein Equivalence Principle; equivalently they are all the *metric* theories of gravity. See (Will, 1993a) for details, and (Staley, 2008, 2013) for a discussion of the empirical grounds for the assumption of Einstein Equivalence in this context.

GTR, all of the rivals to GTR that physicists have regarded as serious, and (infinitely) many other rivals that no one has explicitly entertained, but that are mathematically represented within the framework. Whether a similar error-statistical approach can (or should) be applied to theories regarding aspects of nature other than gravitational physics (particularly in domains such as the biological or social sciences), remains an open question.

However defenders of an error-statistical approach in philosophy of science grapple with these issues, we can see that the frequentist tradition in statistics has given rise to an interesting alternative to the Bayesian approach that challenges long-standing conceptions of the task facing philosophers of science.

10 Realism and anti-realism

10.1 Introduction: success, change, and empiricism

Scientists pose questions about the world and, in some sense, they succeed in arriving, however tentatively, at answers. In this chapter we ask how we should understand that success. In particular, should we think of scientific inquiry as a project aimed at revealing a *true* description of the world, even when that description involves things that we cannot possibly see, hear, touch, taste, or smell? Or should we consider scientists to be engaged in a search for theories that are simply in some way *useful*?

On the one hand, the ways in which scientists are sometimes successful would be, some philosophers argue, inexplicable or miraculous unless it were the case that they have accepted some theories that are (at least approximately) true. The approximate truth of some of their theories is the *best explanation* of certain successes that scientists have enjoyed. *Scientific realism* maintains that we should regard empirically successful theories in the *mature* sciences (like physics, chemistry, and biology) as at least approximately true *because* this is the best explanation of the success of those theories.

On the other hand, we have seen how theories can change in science in dramatic and unexpected ways. A theory might even seem to enjoy tremendous success, only to be replaced by a theory with a radically different conception of the basic entities responsible for the phenomena in its domain. A historically informed caution seems then appropriate with regard to successful theories, calling into doubt the close connection between success and truth that scientific realism postulates. *Theory change* poses a *historical* challenge to scientific realism.

Another criticism of scientific realism alleges it to be inconsistent with a thorough-going scientific attitude. Such an attitude, according to this challenge, should consider the ultimate arbiter of questions of truth to be nothing

other than the observable data. But, the argument goes, theories that involve terms referring to unobservable entities such as electrons necessarily involve statements that go beyond what can be tested with the observable data. So we should understand any commitment scientists make to such theories not in terms of belief in the truth of those theories, but as a pragmatic decision. This is the *philosophical* challenge of *empiricism*.

10.1.1 Example: the hole theory and anti-matter

In 1928, the physicist Paul Dirac published a theory that for the first time described the quantum mechanics of electrons in a way that was consistent with Albert Einstein's Special Relativity (Dirac, 1928). Dirac's theory, however, included something peculiar: electrons with *negative* energy. Such things had never, of course, been observed in any experiment.[1]

Dirac valued the mathematical elegance of a theory very highly, and his own theory exemplified this attribute. So he sought to convert the oddity of negative energy states into a virtue by using them to explain already known particles. This interpretation became known as the 'hole theory.' According to hole theory, there really is an ever-present 'sea' of negative energy electrons. Its density is actually infinite, but because it is uniform the net electromagnetic field from it will be zero everywhere except where there are local divergences from this uniform density. The existence of this sea then helps to ensure that positive energy electrons will stay positive. According to a principle of quantum physics called the exclusion principle, no two electrons can occupy exactly the same state. Since all negative energy states are already occupied, electrons in positive energy states cannot jump into them.

In an effort to unpack all of the consequences of his theory and solve problems threatening it, Dirac then asks what it would be like if there were a *hole* in this sea of negative energy electrons, a location at which no negative energy electron were to be found. His surprising answer is that such a hole would behave just like a particle, but one with an electric charge opposite

[1] Theoretically, negative energy electron states were not entirely new. The classical theory of the electron also included negative energy states, but physicists ignored them as 'unphysical.' This was not possible in a *quantum* theory, because for any electron in a positive energy state, there would be a non-zero probability of that electron undergoing a transition to a negative energy state.

that of electrons. At first he concluded that his theory had just predicted protons, but Hermann Weyl showed that if that were the case protons would have to have the same mass as electrons, which they do not. So Dirac took the radical step of proposing an entirely new kind of particle. He made the prediction with great diffidence:

> A hole, if there were one, would be a new kind of particle, unknown to experimental physics, having the same mass and opposite charge to an electron. We may call such a particle an anti-electron. We should not expect to find any of them in nature, on account of their rapid rate of recombination with electrons, but if they could be produced experimentally in high vacuum they would be quite stable and amenable to observation. (Dirac, 1931, 61)

Meanwhile, a young physicist at Cal Tech named Carl Anderson was beginning a new series of experiments on cosmic rays in a strong magnetic field, using an improved version of the *cloud chamber*, a device that allows physicists to take pictures of the tracks that charged particles leave in a vapor. Anderson, who had no knowledge of Dirac's prediction, noticed in late 1931 that some of his photographs seemed to depict positively charged particles with masses considerably less than that of a proton. They had to be charged because their paths curved (as particles carrying a charge should do when passing through a magnetic field), and the charge had to be positive. (Anderson had inserted a lead plate in his cloud chamber, allowing him to determine whether the particles entered the chamber from below or from above. He could then infer the sign of the charge by noting whether the track curved to the left or the right.) Moreover, the particles could not be protons, but had to have masses similar to that of an electron, because of the very thin tracks they left. A proton, with mass about 1800 times that of an electron, would leave a much thicker track. Based on a number of photographs such as the one in Figure 10.1 (C. Anderson, 1933, 492), Anderson proclaimed the discovery of the "positive electron," which he dubbed the "positron" (C. Anderson, 1933).[2]

[2] Anderson had set out not to find new particles but to investigate the origins of cosmic rays, a question of great interest to the head of Cal Tech, Robert A. Millikan (C. Anderson, 1999). C. T. R. Wilson had originally developed the cloud chamber in order to investigate the formation of clouds, not to study cosmic rays or discover new particles (Chaloner, 1997; Galison & Assmus, 1989). The unfolding of scientific discovery cares little for the expectations of those who pursue it!

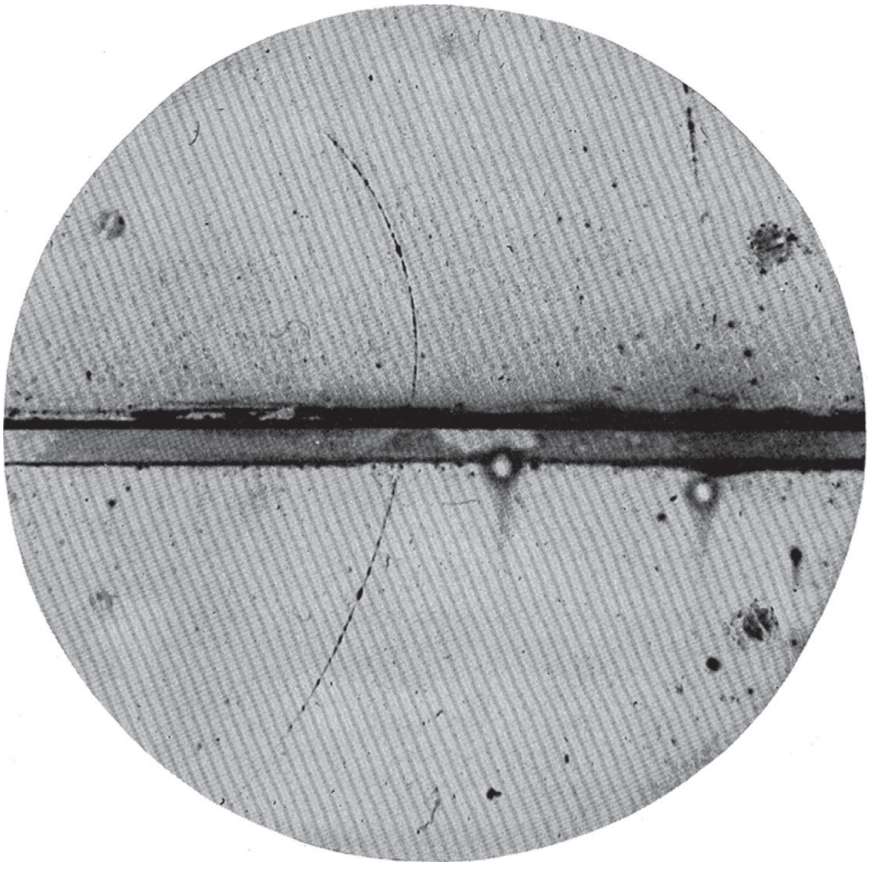


Figure 10.1 A cloud chamber photograph of a track left by a positron. The particle enters from below and passes through a 6 mm lead plate across the center. The subsequent loss of momentum results in the greater curvature of the track in the upper region. (Reprinted with permission from Carl Anderson, "The positive electron." *Physical Review* 43, 492 (1933). Copyright 1933 by the American Physical Society.)

Physicists expressed skepticism at first, but more detailed experiments by P. M. S. Blackett and G. P. S. Occhialini provided more detailed measurements. Unlike Anderson, Blackett and Occhialini did discuss their findings in light of Dirac's theory, concluding that although their data were not sufficient to test it *directly*, Dirac's theory "predicts a time of life for the positive electron that is long enough for it to be observed in the cloud chamber but short enough to explain why it had not been discovered by other methods" (Blackett & Occhialini, 1933, 716).

Further experimentation strengthened the positron's status as an elementary particle and as the first of a whole class of anti-particles. Dirac had effectively predicted this, too. In the same paper he argued that the proton – the only other particle known at the time – would also have a sea of negative energy states, holes in which would constitute anti-protons. In spite of its seemingly odd way of viewing the world, Dirac's theory was strikingly successful.

10.2 'No miracles': an argument for realism

When a theory is able to get things right in this way, it seems intuitively as though the scientists who developed it must be onto something. That intuition lies at the heart of one argument for *scientific realism*. This argument (sometimes called the 'no miracles argument') asks *what is the best explanation* of the fact that certain scientific theories provide a basis for scientists to predict successfully the outcomes of experiments never before performed and for engineers to extend our ability to intervene in and control natural processes? Either Dirac got tremendously, even miraculously, *lucky*, or else he had come up with a theory that was, if not *exactly* true, at least close to the truth. As stated in an often-quoted phrase from Hilary Putnam, "realism is the only philosophy that does not make the success of science a *miracle*" (Putnam, 1979, 73; emphasis in original).

Although statements of the realist thesis about scientific theories vary somewhat, the basic idea is this:

Thesis 2 (Realism) *Scientific theories that achieve a certain level of success in prediction and experimental testing are (probably) approximately true.*

What is the alternative? Anti-realists regard theories, even successful ones, as somehow able to perform their function without being true, or at least without being literally true. One version of this idea is *instrumentalism*, which holds that theories should be regarded as tools (or instruments) for making predictions and exerting control over natural processes. As such, the instrumentalist insists, theories can serve their function without being even approximately true. Thus, from the fact that a theory serves its function successfully, we should not infer that it is even approximately true.

Advocates of scientific realism contend that the argument they offer is of the same sort that scientists use to defend their theories: they consider

scientific realism to be superior to alternative ways of thinking about theories such as instrumentalism because it provides the best explanation of the success of certain scientific theories. Because, they argue, scientists themselves accept, in any given domain, the theory that offers the best explanation of the phenomena in that domain, scientific realism earns the appellation 'scientific' not simply because it is a thesis about science, but because it uses a scientific argument in defense of that thesis.

Scientific realists thus adopt a model of the reasoning by which scientists defend scientific theories called *Inference to the Best Explanation* (IBE). Another term sometimes used for this kind of inference is *abduction*.[3] To infer to the best explanation is to infer "from the premise that a given hypothesis would provide a 'better' explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true" (Harman, 1965, 89). Although philosophers disagree over which features make an explanation the *best*, some prominent considerations have included plausibility, consilience, simplicity, and the extent to which a proposed explanation is analogous to explanations already known to be true (Lipton, 2004; Thagard, 1978).

More specifically, the argument for scientific realism focuses on the ways in which scientists rely on methods that depend on theoretical assumptions. Such assumptions, they argue, enter into the design of experiments, the analysis of data, the use of instruments, and the evaluation of evidence. "All aspects of scientific methodology are deeply theory-informed and theory-laden" (Psillos, 1999, 78). Moreover, scientists succeed in conducting experiments and making accurate predictions while relying on these theory-laden methods. The best explanation for such success is that "the theoretical statements which assert the specific causal connections or mechanisms by virtue of which scientific methods yield successful predictions are approximately true" (Psillos, 1999, 78; see also Boyd, 1989).

10.2.1 Circularity?

Precisely because scientific realism relies on IBE as a mode of argument, critics have alleged that the argument for scientific realism is viciously

[3] Charles S. Peirce introduced the term 'abduction,' but the kind of inference he intended for it to characterize is not quite that which contemporary philosophers typically have in mind when they use the term.

circular. They point out that instrumentalists *reject* IBE as a means of justifying the belief that a theory is true. Were they to accept the IBE principle that the best explanation of the phenomena in a given domain should be regarded as true, they would already be realists about scientific theories.

Since skepticism about IBE is a chief motivation for instrumentalism, the most that an instrumentalist would conclude from the fact that scientific realism is the best explanation of empirical success in the sciences is that scientific realism is instrumentally useful (Fine, 1986; van Fraassen, 1980). (They may not concede even that much. Unlike the scientific theories alluded to in the scientific realist argument, scientific realism itself has not provided us with any successful predictions; nor has it enabled anyone to do a better job of analyzing data, devising instruments, or performing experiments (Laudan, 1981, 46).) Since the argument for scientific realism assumes the reliability of a mode of reasoning that is rejected by those who reject the conclusion of that argument, the defense of scientific realism assumes that which is at issue and hence begs the question.

Defenders of scientific realism such as Stathis Psillos respond to this charge by drawing an important distinction between two different ways in which an argument can be circular. An argument is *viciously* circular, or *premise-circular* (Braithwaite, 1953, 274–285), if it includes among its premises a claim that is identical or equivalent to the conclusion of that argument (and would not be valid without that premise). It is unreasonable to take such an argument as a reason to adopt a belief in its conclusion. If the proposition expressed by the conclusion is not one that you accept, then the argument depends crucially on a premise you do not accept. If it is already among your beliefs, then accepting the conclusion of the argument does not involve adopting any new belief. Note that what is at issue is not the validity of the argument, but whether the argument provides a rational basis for accepting its conclusion.

Defenders of scientific realism then point out that the defense of scientific realism by appeal to IBE is not circular in this way, because it does not include *as a premise* the claim that IBE is reliable.

That distinction matters because although it is true that a person must accept the premises of an argument for that argument to be rationally persuasive, it is not (or at least not obviously) true that for an argument to be rationally persuasive a person must believe that the kind of argument used is reliable. It is only necessary that the form of argument used *is* reliable. So scientific realists can argue cogently for their position without first

convincing those who are skeptical of their position to accept that IBE is reliable, provided that IBE is *in fact* reliable.

You may be thinking "that seems like a bit of a dodge," but recall from Chapter 1 how defenders of induction run into a similar difficulty. Justifying the use of induction in an inference requires another inductive argument.[4] Even deductive inferences are subject to this difficulty. Could you imagine defending the reliability of deductive reasoning without using deductive reasoning? So long as deductive reasoning is *in fact* reliable (in the sense of being truth-preserving), this need not lead to a debilitating skepticism. So in this respect, abductive reasoning such as scientific realists employ is no worse off than inductive or deductive inference (Psillos, 1999, 89).

We should approach this 'no worse off' reply carefully, however. It shows that the form in which this circularity arises should not *by itself* cause us to reject the kind of reasoning being used, as that would lead to the self-defeating conclusion that we should also reject deductive and inductive inferences. That might defeat any attempt to reject IBE as unreliable, were such skepticism based *only* on the charge of circularity against attempts to justify IBE. Those skeptical of IBE, to be taken seriously, need also to provide some kind of evidence that IBE is *not actually* reliable.

Can critics of scientific realism produce such evidence? Let's consider the historical record.

10.3 The problem of theory change

Scientific theories have come and gone over the years. Newton's emission theory of light enjoyed quite a few years of success before nineteenth-century physicists abandoned it in favor of the theory that light consists of waves in an aether filling all of space. The idea of the aether as a mechanical medium for light vibrations proved fruitful for a time, but was found ultimately unworkable, while the electromagnetic theory of James Clerk Maxwell provided an alternative conception that identified the waves with fluctuations in an electromagnetic field. In light of twentieth-century developments in quantum theory, physicists now decline to regard light as *fundamentally* either a wave or a particle.

[4] Gilbert Harman argues that all inductive inferences really are inferences to the best explanation (Harman, 1965). For purposes of the present argument, I will assume that induction and IBE are distinct.

In an influential 1981 essay, Larry Laudan argues that such examples of successful theories later rejected in favor of conceptually quite distinct successors, pose a significant challenge to scientific realism by calling into question the connection between the truth or even approximate truth of a theory and its empirical success.

According to Laudan, the scientific realist is committed to the claims that (1) "Scientific theories (at least in the 'mature' sciences) are typically approximately true and more recent theories are closer to the truth than older theories in the same domain" and that (2) "The observational and theoretical terms within the theories of a mature science genuinely refer." Laudan aims his arguments at a "convergent" version of scientific realism committed to the further claims that (3) "Successive theories in any mature science will be such that they 'preserve' the theoretical relations and the apparent referents of earlier theories" and (4) "Acceptable new theories do and should explain why their predecessors were successful insofar as they were successful" (Laudan, 1981, 20–21).

The scientific realist argues that we should accept these claims because they explain the empirical success of the scientific theories in question. For such an explanation to succeed, Laudan argues, the following two principles must be true:

Principle 6 (Reference connection) *If the central terms in scientific theories genuinely refer, those theories will generally be empirically successful.*

Principle 7 (Truth connection) *If scientific theories are approximately true, they will typically be empirically successful.*

Laudan offers counterarguments to these principles on several fronts, drawing on the history of science. For example, he argues against Principle 6 by first distinguishing two claims that might provide support for it:

Principle 8 (Reference to success) *A theory whose central terms genuinely refer will be a successful theory.*

Principle 9 (Success to reference) *If a theory is successful, we can reasonably infer that its central terms genuinely refer.*

Laudan's attack on Principle 8 points to theories with (what we currently regard as) genuinely referring terms that were *not* empirically successful (such as the original version of tectonic plate theory, as articulated by Alfred Wegener). Principle 9 must contend with historical examples of theories

whose central terms we currently regard as *non-referring* that nonetheless were empirically successful (such as the phlogiston theory of combustion, or the theory that light is a vibration in the 'luminiferous aether'). Moreover, Laudan points out, Principle 9 is not strong enough by itself to support the scientific realist's contention that scientific realism *explains* the success of science, since it tells us not that we can expect genuinely referential theories to be successful, but only that we can expect successful theories to be genuinely referential. The scientific realist might be committed to the latter claim, but the *argument for* scientific realism depends on the former claim.

The argument Laudan offers against Principle 7 builds on these points. Parallel to the previous distinction with regard to reference, we can point to two commitments of the scientific realist with regard to truth:

Principle 10 (Approximate truth to success) *If a theory is approximately true, then it will be empirically successful.*

Principle 11 (Success to approximate truth) *If a theory is empirically successful, then it is probably approximately true.*

Laudan acknowledges that Principle 10 enjoys a certain air of plausibility, but that, he claims, is due to its resemblance to the quite distinct and "self-evident"

Principle 12 (Truth to success) *If a theory is true, then it will be successful.*

Of course if a theory is true, we should expect it to succeed, but scientific realists cannot take much comfort from this, because their thesis does not assert only that being *true* explains the success of theories, but that being *approximately true* does so. Absent a theory of approximate truth that would underwrite Principle 10, the realist can issue only a promissory note here.

Moreover, the history of science provides ample reasons to reject Principle 11, provided that theories postulating non-existent entities are not approximately true. Laudan expands on his earlier discussion of theories with non-referential terms that were empirically successful. His long list of examples includes the "crystalline spheres of ancient and medieval astronomy," "the effluvial theory of static electricity," "the caloric theory of heat," "the vital force theories of physiology," and the electromagnetic and optical aethers.

Indeed, as Thomas Pashby has argued (Pashby, 2012), Laudan might have added Dirac's 'sea' of negative energy electrons to this list. Although

Dirac's theory enjoyed genuine empirical success, including its remarkable prediction of positrons, his proposal for understanding the theory did not survive. Through the middle decades of the twentieth century, physicists such as Sin-Itiro Tomonaga, Julian Schwinger, Freeman Dyson, and Richard Feynman developed mathematical techniques to handle the problem of infinities in Dirac's theory (Schweber, 1994). The result was an approach to quantum electrodynamics that describes particles and anti-particles in terms of fields rather than seas of negative energy particles.

10.3.1 The pessimistic meta-induction

The list of successful, not-even-approximately-true theories can, Laudan asserts, "be extended ad nauseam" (Laudan, 1981, 33) and is so extensive that writers have taken it as the basis for the *pessimistic meta-induction*. Authors have used this term to refer to two distinct arguments. The first, due to Hilary Putnam, argues (in a *reductio ad absurdum*) that if we assume our current theories are true, then we must regard past theories as false, which provides us with an inductive reason to expect our current theories to turn out also to be false (Putnam, 1978). We will not concern ourselves further with this version of the pessimistic meta-induction. The inductive argument on which it relies is weak at best, and Putnam himself may not have taken it seriously (P. J. Lewis, 2001).

Like Putnam's, Laudan's argument begins by assuming, as the realist believes, that most currently successful theories in the mature sciences are approximately true. But from this his argument seeks to show not (directly) that we should believe our current theories will turn out to be false, but only that the empirical success of a theory is not a reliable indicator of its truth or even its approximate truth. Even accepting Laudan's historical claims about the prevalence of empirically successful theories that are not even approximately true by the light of currently accepted science, it remains unclear what follows from this regarding the relation between success and truth. For example, suppose we think of empirical success as a kind of diagnostic test for truth. The fact that it has often given a positive result for false theories does not by itself prove that it is unreliable. Suppose it is reliable in the sense that it rarely yields a negative result when applied to a true theory (the *false negative rate* is low) and rarely yields a positive result when applied to a false theory (the *false positive rate* is low). We might still see numerous examples

of positive results from this test as applied to false theories, if most of the theories to which it has been applied are false. As Peter Lewis has pointed out, this leaves a way for the realist to explain Laudan's historical evidence: the population of *past* theories to which the realist's test for truth has been applied consists of mostly false theories, thus yielding a significant number of false positives. But the kind of realist Laudan is criticizing believes scientific progress entails that we now have many more true theories than we have had in the past (P. J. Lewis, 2001; see also Magnus & Callender, 2004).

Applying such statistical ideas to the history of science is problematic on a number of levels. Do theories form a *population*, from which scientists have sampled? Does the realist employ success as a diagnostic test to be applied to such a population? One sure conclusion is that historical change complicates any attempt to reason inductively from the historical record.

Nonetheless, Laudan's argument poses at least this challenge to the realist: What supports the claim that IBE is a reliable form of argument, given that it has so often misled us in the past? The rule-circularity of the IBE argument for scientific realism may not by itself lead us to doubt its cogency, but the historical record of apparently strong but erroneous IBEs suggests that we are owed an argument in support of the reliability of IBE.

10.4 The empiricist challenge

The other significant challenge that scientific realism has confronted has its source not in the history of science but in a philosophical orientation historically associated with a scientific outlook: *empiricism*. Empiricist philosophers share a commitment to accounting for human knowledge in terms of experience gained through the use of our senses.[5] Empiricist critics of scientific realism argue that believing in the truth of a theory that refers to unobservable things goes beyond what experience can warrant.

The logical empiricists provide one example (see Chapter 4). In their attempt to distinguish between cognitively meaningful discourse and meaningless metaphysics, the logical empiricists sought to characterize the responsible use of theoretical language in terms of the use of *correspondence*

[5] This statement suffers from an unavoidable vagueness. Empiricists must regard the nature of experience and its role in knowledge production as empirical questions that cannot be settled at the outset by philosophical fiat. Bas van Fraassen concludes that empiricism cannot be characterized by a belief in any particular thesis, but must be understood as a *stance* or orientation in one's theorizing (van Fraassen, 2002).

rules that would link theoretical terms (like 'electron') to terms describing observable things (like tracks in a cloud chamber photograph). The correspondence rules were to provide the *meaning* of the theoretical terms. So a physicist who avows belief in a theory about electrons would not need to believe in the existence of subatomic particles zooming around with their miniscule mass and electric charge; she would commit herself instead to the existence of a theory in which expressions using such terms are deductively linked, via correspondence rules, to descriptions of observations and experimental outcomes.

10.4.1 Constructive empiricism

In his 1980 book *The Scientific Image* Bas van Fraassen set forth such a bracing challenge to realists that it established much of the framework in which subsequent discussion of the issue has taken place. Van Fraassen advocates a position called *constructive empiricism*:

Thesis 3 (Constructive Empiricism) *Science aims to give us theories that are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate.* (van Fraassen, 1980, 12)

Van Fraassen's rough account of empirical adequacy states that a theory is empirically adequate "exactly if what it says about the observable things and events in this world, is true" (van Fraassen, 1980, 12). For greater precision, recall the contrast in Chapter 4 between syntactic and semantic views of scientific theories. Van Fraassen endorses a semantic view. For him, to give a scientific theory is to specify a family of structures: the models of the theory. Such a specification will include the *empirical substructures* of those models, which potentially represent observable things and events. An empirically adequate theory, then, is one for which all of the data ("the structures which can be described in experimental and measurement reports," or what van Fraassen calls *appearances*) correspond to elements of the empirical substructure (van Fraassen, 1980, 64). By "all the data" van Fraassen means *all* the data: not just all observations recorded so far, but all those that will be in the future.[6] To state it in more intuitive but less precise terms, an empirically

[6] Of course, the data might have noise in them. Moreover, some data may be erroneous due to instrumental flaws or errors in experimental design or execution. A more careful statement of the definition of empirical adequacy would need to take account of these realities of data collection.

adequate theory is a collection of models into which all of the appearances fit.

Unlike the logical empiricist, the constructive empiricist does not seek to interpret away a theory's apparent reference to unobservable things. If a theory uses language that refers to electrons and is true, then it entails the existence of electrons. However, the constructive empiricist regards such a theory as contributing to science, not by providing a true description of an unobservable micro-world, but rather by describing a set of models into which all of the data might possibly fit. Those models might also refer to things and events that do not correspond to any possible data, but the acceptance of a theory does not involve belief in *those* things, only in the claim that the theory is empirically adequate, i.e., that the data will fit into its empirical substructure.

Thesis 3 makes no descriptive claim about the beliefs of scientists (perhaps most of them are realists), but a *prescriptive* claim about the aims of scientific inquiry. It tells us what *should* count as a successful contribution to science. Note also that to say that a theory is empirically adequate requires an inductive leap beyond what the current data state, for it entails that the theory will be consistent with all *future* data.

10.4.2 What is observable?

Because van Fraassen defines empirical adequacy in terms of what is observable, the exact meaning of the constructive empiricist thesis depends on what it means to be observable. Two questions confront the constructive empiricist: (1) Can one state a consistent and defensible criterion of observability? and (2) Does observability *matter* in a way that would justify giving it such a central role in our understanding of the aims of science?

Beginning with the first question, van Fraassen tackles an argument due to Grover Maxwell against the possibility of a principled distinction between what we can and cannot observe. Maxwell argues that acts of observation constitute a continuum from the direct and "unaided" observation of an object, through observation with such simple aids as a magnifying glass or a low-power optical telescope or microscope, to cases involving more technological mediation, such as the use of radio telescopes or electron microscopes or bubble chambers. This continuum leaves us, according to Maxwell "without criteria which would enable us to draw a non-arbitrary line between 'observation' and 'theory' " (Maxwell, 1962, 7).

In response, van Fraassen proposes the following criterion of observability, commenting that "This is not meant as a definition, but only as a rough guide to the avoidance of fallacies" (van Fraassen, 1980, 16):

Principle 13 *X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it.*

Maxwell's argument indicates that we can draw no sharp line along the continuum separating the certainly and unequivocally observable from the certainly and unequivocally unobservable. But this only indicates that 'observable,' like many other perfectly meaningful and useful words, is a *vague predicate*. By pointing to clear cases and counter-cases, we can provide evidence for the meaningfulness of such predicates – like 'bald', 'blue', or 'bilingual' – even if in other cases we are not sure whether they apply or not. The moons of Jupiter are observable because we not only can observe them through the telescope but "astronauts will no doubt be able to see them as well from close up" (van Fraassen, 1980, 16). A subatomic particle passing through a bubble chamber is clearly not observable. You may see the track of bubbles that it leaves, but under no condition can you observe the particle itself.

Such a response suffices to establish that the distinction between what is and is not observable is meaningful, but leaves our second question, which concerns, in Maxwell's words, "what ontological ice" such a distinction cuts (Maxwell, 1962, 8). Van Fraassen tackles one aspect of this problem raised, again, by Maxwell, as a problem concerning the relevant range of possible circumstances in which we might observe a thing. For example a mutation that enabled a person to form visual images using radiation of wavelengths different from those that human vision normally utilizes would allow one to observe many things we cannot currently observe (Maxwell, 1962, 11). This suggests that the delineation of the observable from the unobservable is *indeterminate*, since we cannot predict just what possible modes of sensory experience we might develop in the future.

Van Fraassen's reply is to insist that we should understand the possibilities of observation in terms of actual, present human capabilities: "The human organism is, from the point of view of physics, a certain kind of measuring apparatus. As such it has certain inherent limitations … It is these limitations to which the 'able' in 'observable' refers — our limitations, *qua* human beings" (van Fraassen, 1980, 17). Constructive empiricism thus pins the aims

of science to something of a moving target. To the extent that our cognitive capacities change, so will the meaning of empirical adequacy. This, however, van Fraassen accepts as an entirely reasonable commitment, for the alternative is to require that "our epistemic policies should give the same results independent of our beliefs about the range of evidence accessible to us" (van Fraassen, 1980, 18).

Does that answer our second question, though? We want to know why observability should make a difference to what we should believe regarding an empirically successful theory.

10.4.3 Underdetermination

Constructive empiricism is grounded in a certain view about what empiricism requires: only the appearances should matter to the epistemic commitments of science. Since theories with different non-empirical substructures can share the same empirical substructures, a decision to believe one such theory rather than another would have to attribute epistemic import to such *non-empirical* virtues as simplicity (in one or more of its guises), mathematical elegance, fruitfulness, and so on. Van Fraassen does not deny that scientists value such features in a theory, but insists that an empiricist must value them only as pragmatically *useful*, not as signs of truth. Dirac may have considered his theory of the electron to be mathematically beautiful, but to the extent that he took that aesthetic value as a reason to believe his theory true (as opposed to making a pragmatic commitment to develop and defend it), he would be adopting an epistemic attitude based on something other than the appearances. A consistent empiricist would not believe in Dirac's 'sea' of negative energy electrons, or in holes in that sea, which after all are absent from later formulations of quantum theories of electrons. Nor should one believe in electrons and positrons themselves, which some future theory making all the same experimental and observational predictions might eliminate.

Constructive empiricism therefore relies strongly on an appeal to a kind of *underdetermination*. That a currently embraced theory that refers to unobservables gives the best explanation of the phenomena – even if it does so with respect to all *future* phenomena – is no warrant for the truth of the theory because of other empirically equivalent theories that differ only with respect to theoretical virtues that an empiricist cannot take as a reason to believe

those theories that possess them. One might, however, resist this argument if one could make the case that such theoretical virtues are in fact reliable indicators of the truth of a theory (McAllister, 1999; Musgrave, 1982).

Van Fraassen employs a different underdetermination argument to defeat one strategy for defending scientific realism. Scientific realists sometimes argue that scientists should use Inference to the Best Explanation as a means of inferring the *truth* of scientific hypotheses because we do this in ordinary circumstances and we should consistently apply the same kinds of reasoning in both ordinary circumstances and in science. To adapt van Fraassen's own example: Zig has left an (obviously) unwanted horseradish doughnut on the kitchen counter overnight. When he wakes, he finds no doughnut, but observes in its place a few crumbs, and some small black pellets. Zig infers H_M: a mouse has taken up residence (and has a higher opinion of horseradish doughnuts than does Zig). Of course, he has not seen the mouse, but the best explanation of the evidence that he has seen is that H_M is true. It would be silly to insist that Zig infer only that things are *as if* there were a mouse, but to remain agnostic about the existence of an actual mouse. Likewise, it would be silly to ask the scientist to remain agnostic about positrons in the face of the evidence compiled by Anderson and by Blackett and Occhialini.

This argument interprets Zig's reasoning as an instance of IBE in which Zig infers the truth of the hypothesis. But van Fraassen points out that one could equally well describe Zig's inference regarding the mouse as an inference to the empirical adequacy of H_M. Mice being observable, H_M's truth and its empirical adequacy amount to the same thing. All of its consequences are observable, so it is true exactly if it is empirically adequate, and vice versa. Van Fraassen thus offers a rival hypothesis to that of the scientific realist regarding how we reason: "we are always willing to believe that the theory which best explains the evidence is empirically adequate." This hypothesis, he claims, "can certainly account for the many instances in which a scientist appears to argue for the acceptance of a theory or hypothesis, on the basis of its explanatory success" (van Fraassen, 1980, 20). The empirical evidence of how people do in fact reason underdetermines the choice between realist and anti-realist philosophies of science.

We turn now to some of the ways realists and anti-realists have elaborated or modified their positions and arguments over years of debate. Various forms of realism and arguments for it have emerged from these efforts.

10.5 Strengthening scientific realism

In the face of Laudan's historical challenge, scientific realists have tightened their argument by both refining their standards for what counts as success, and by distinguishing those aspects of the theory that must be regarded as approximately true to explain such success. In this way, scientific realists have refined their IBE argument by modifying their explanation. They have modified both the *explanans* (the propositions doing the explaining) and the *explanandum* (the proposition describing what is being explained).

10.5.1 Maturity and success

We can start with their modification of the idea of success. Laudan's argument appeals to a long list of theories that enjoyed periods of empirical success, although we now regard them as false and their central terms as non-referential. In defense, scientific realists have sought to strike items from this list by invoking a more stringent standard of empirical success than Laudan's.

One such standard requires theories to yield successful *novel predictions* to qualify for a realist attitude (Leplin, 1997; Psillos, 1999). In Chapter 6 we discussed Lakatos's use of a novel prediction criterion to distinguish progressive from degenerating research programs. Philosophers have variously interpreted this standard as requiring a theory to predict some new result that was not known before, or is not predicted by any of the theory's rivals, or would be very surprising or unexpected independently of that theory, or was not used in the construction of that theory. In his defense of scientific realism, Stathis Psillos employs the concept of *use novelty* advocated by John Worrall (Psillos, 1999; Worrall, 1985, 1989a), whereas Jarrett Leplin defends realism with an analysis of novelty that combines and refines elements of both Worrall's and Lakatos's conceptions (Leplin, 1997).

However conceived, the novel prediction standard significantly reduces the number of episodes in the history of science in which a successful theory turned out to be incompatible with our current best theories. Psillos argues in addition that many of Laudan's successful-but-false theories emerged at times when their respective disciplines were not yet *mature*. They had not yet produced "a body of well-entrenched beliefs about the domain of inquiry which, in effect, delineate the boundaries of that domain, inform theoretical research and constrain the proposal of theories and hypotheses"

(Psillos, 1999, 107). By thus invoking higher standards, scientific realists can thus escape any threat from such putative successes as the humoral theory of medicine, the crystalline spheres theory of medieval astronomy, or the effluvial theory of static electricity.

10.5.2 Divide and conquer

Raising the standards for maturity and success does not, however, suffice to address the entirety of Laudan's challenge, for some significant episodes remain in which a genuinely mature science seems to have succeeded in producing novel successes from a theory with non-referring terms. Both realists and anti-realists repeatedly turn their attention to the case of the nineteenth-century wave theory of light and the luminiferous or optical aether, which physicists used to derive some strikingly successful novel predictions, although the latter plays no role in our current understanding of optical phenomena.

Psillos summarizes the strategy for answering this challenge: "it is enough [to defeat Laudan's argument] to show that the success of past theories did not depend on what we now believe to be fundamentally flawed theoretical claims" (Psillos, 1999, 108). He calls this strategy *divide et impera* (divide and conquer). The strategy has important consequences for the *methods* used by defenders of scientific realism and for the *content* of the scientific realist position.

The divide and conquer strategy requires defenders of scientific realism to conduct historical investigations to determine just what theoretical claims *were* essential to a given successful prediction, and how scientists evaluated the contributions of various elements of the theory to such a prediction. Psillos, for example, argues at length, based on his reading of the contributions of various nineteenth-century physicists, that we should distinguish between their mathematical description of the dynamics of light propagation, the essentials of which carried over from wave optics into the later reconceptualization of light in terms of the electromagnetic field, and their attempts to construct a mechanical model of an elastic aether, the vibrations of which could be understood as light waves. Whereas the former played an important *explanatory* role (and were preserved in later theories), the latter were merely *illustrative*, and later theories of light excluded them precisely because physicists could not give them a description from which

they could predict the behavior of light. He summarizes the situation thus: "The parts of 'luminiferous aether' theories which were taken by scientists to be well-supported by the evidence and to contribute to well-founded explanations of the phenomena were retained in subsequent theories. What became paradigmatically abandoned was a series of models which were used as heuristic devices for the possible constitution of the carrier of light-waves" (Psillos, 1999, 140).

In this way, scientific realists seek to turn the problematic case of the luminiferous aether into a success story for a more nuanced realist position. The case of Dirac's sea of negative energy electrons might pose a more difficult challenge. Thomas Pashby has recently argued that the early successes of Dirac's theory depended crucially on the sea and could not have been derived without it (Pashby, 2012).

The divide and conquer strategy furthermore requires the scientific realist to adopt a more differentiated attitude toward successful theories than Thesis 2 would suggest. We should not regard a theory in its entirety as 'approximately true' simply because that theory successfully predicts novel phenomena; we should adopt a realist attitude only toward those "constituents which contribute to successes and which can, therefore, be used to account for these successes" (Psillos, 1999, 110).

10.6 Experimental routes to realism

Thus far, our discussion of realism and anti-realism has focused on the interpretation of *theories* that provide explanations making reference to unobservable entities. Debates over realism, thus construed, partake of a preoccupation with theories and their status typical of most work done in philosophy of science up until the early 1980s. Then, a few philosophers began to argue that such 'theory-dominated' philosophy of science resulted in a distorted understanding of scientific knowledge that privileged abstract logical or quasi-logical relations between theory and evidence over material engagement and interaction with experimentally induced phenomena. These "New Experimentalists" (Ackermann, 1989) argued that valuable philosophical lessons were to be learned by attending to the practices of experimentation. Scientists, they insisted, do not only record observations so they can formulate observation statements to put into some formal relationship with theoretical statements. They intervene in the world,

tinkering with technological systems to *make* them convey information about the phenomena that interest them (Ackermann, 1985, 1989; Bogen & Woodward, 1988; Franklin, 1986; Galison, 1987; Gooding, 1990; Hacking, 1981, 1982, 1983; Hon, 1987, 1989; Mayo, 1996). In the words of Ian Hacking, "Experimentation has a life of its own" (Hacking, 1983, 150).

10.6.1 Entity realism

Ian Hacking made some of the earliest and most influential of contributions to the New Experimentalism. Hacking argued that the basis for a realist attitude about unobservable entities such as electrons lay not in the fact that our explanatory theories make apparent reference to them, but in the fact that we can *use* them to make things happen – especially in the course of investigating things other than electrons. Hacking described an experiment that used an "electron gun" to investigate the phenomenon of parity violation in what physicists call "weak neutral currents." We will pass over an explanation of the physics of weak neutral currents here – Hacking himself gives only the briefest discussion – because the crucial aspect of the experiment for Hacking's point is not the physics being investigated but the *engineering* enterprise of building an electron gun that makes an experimental investigation of that physics possible. He summarizes his point thus: "We are completely convinced of the reality of electrons when we regularly set out to build – and often enough succeed in building – new kinds of devices that use various well understood causal properties of electrons to interfere in other more hypothetical parts of nature" (Hacking, 1983, 77). Elsewhere, referring to an experiment in which physicists altered the charge on a sphere of niobium by "spraying" it with positrons, Hacking quips: "So far as I'm concerned, if you can spray them, then they are real" (Hacking, 1983, 23).

Hacking does not simply observe that experimenters as a matter of fact tend to believe in the existence of entities that they manipulate to investigate phenomena. Instead he claims that the "enterprise" of using entities to experimentally investigate the world "would be incoherent without" a realist attitude toward the entities thus manipulated. Given that the experimental enterprise "persistently creates new phenomena that become regular technology," this enterprise is evidently not incoherent (Hacking, 1982, 73). Thus, "engineering, not theorizing, is the proof of scientific realism about entities" (Hacking, 1982, 86).

The realism Hacking endorses *only* concerns entities that experimentalists use to learn about other phenomena. He does not endorse realism about theories as such, and denies that his argument relies on the kind of inference to the best explanation argument that scientific realists typically employ. (At around the same time, Nancy Cartwright arrived at a similar view, following a different path. In *How the laws of physics lie*, she argues that the laws of physics have great explanatory power precisely because they do *not* describe actually occurring states of affairs ("the facts"). But, she insists, the laws of physics do truthfully describe *something*: they "describe the causal powers that bodies have" (Cartwright, 1983, 61).)

Hacking's writings on entity realism seem to point to two distinct arguments with different conclusions, but Hacking so intermingles them that he may not see them as distinct. One argument seeks to establish good evidence for the claim that not only do unobservable entities exist, but we make regular and reliable use of them in our experimental undertakings: "The 'direct' proof of electrons and the like is our ability to manipulate them using well understood low-level causal properties" (Hacking, 1982, 86). Other passages, however, point to a different argument. When Hacking declares that the fact that "experimenters are realists about the entities that they use in order to investigate other hypotheses or hypothetical entities" is "not a sociological fact" but a requirement of the evident *coherence* of the enterprise of scientific experimentation (Hacking, 1982, 73), his argument is not a "direct" proof of the existence of electrons or anything else. It does not establish the truth of any scientific realist claim. It argues that experimental scientists should adopt realist *beliefs* about certain kinds of entities because doing so contributes to the coherence of their enterprise. A belief can contribute to the coherence of something without being true. The child who believes that the crumbs left on the plate by the Christmas tree are the consequence of Santa Claus's midnight snack contributes to the coherence of the story her parents tell her about the presents that have appeared overnight.

Independence from theory?

The more philosophically substantial of Hacking's arguments, then, is the first one. David Resnik has reconstructed that argument as follows:

> (1) We are entitled to believe that a theoretical entity is real if and only if we can use that entity to do things to the world. (2) We can use some theoretical entities, e.g. electrons, to do things to the world, e.g. change the charges of

> niobium balls. (3) Hence, we are entitled to believe that some theoretical entities, e.g. electrons, are real. (Resnik, 1994, 401)

Under this reconstruction, critics can either raise objections against premise (1) (e.g., Shapere, 1993) or premise (2). Resnik targets premise (2), not to dispute the soundness of Hacking's argument, but to dispute Hacking's claim that the entity realist need not rely on inference to the best explanation, and thus can maintain an anti-realist attitude to *theories*. Resnik's critique attempts to show that justifying premise (2) requires the entity realist to argue that the best explanation of the reliability of the instruments experimenters use is given by "low-level generalizations (theories?) and other assumptions" that describe those instruments as taking advantage of theoretical entities (Resnik, 1994, 404). In other words, the realist about positrons cannot simply claim without justification to spray them, she must warrant that claim by invoking an explanation of the functioning of her instruments that appeals to at least some theoretical claims.

Michela Massimi also disputes Hacking's claim to divorce entity realism from realism about theories. Her version of the argument does not saddle the theory realist with an appeal to inference to the best explanation, however. Massimi relies on an underdetermination argument to raise a worry for the entity realist that parallels the underdetermination problem for realism about theories: For any given claim about using a particular entity to produce certain effects in an experimental setting, distinct but *empirically equivalent entities* might produce the same effects. Massimi does not raise this possibility to underpin a skeptical argument based on underdetermination. Rather, she invokes experiments in particle physics testing the *quark* theory of subatomic particles against the rival *parton* theory to argue that the very experimental evidence that suffices to resolve underdetermination regarding which entity the experimentalist is using also provides the grounds on which to choose one theoretical model over another. This makes theory-free entity realism an unstable position: one either foregoes appeals to theory, leaving one vulnerable to objections based on empirically equivalent entities, or one embraces theory, committing to realism about at least some theoretical claims in order to vindicate realism about entities (Massimi, 2004).

10.6.2 A direct experimental argument for realism?

Peter Achinstein has advanced another argument in support of realism that appeals to experiment. Drawing upon a suggestion by Wesley Salmon

(Salmon, 1984, 213–227), Achinstein defends the claim that Jean Perrin's experimental investigations of Brownian motion, carried out in 1908, demonstrate scientific realism by giving an experimental demonstration of the existence of molecules.

Brownian motion, which derives its name from the English botanist Robert Brown who observed it in 1827, is the erratic and ongoing movement of microscopic particles suspended in liquids. Perrin carried out a variety of experiments in which he observed through a microscope the motion of particles of gamboge (a resin derived from certain trees) in an emulsion. Perrin derived an equation relating observable quantities by assuming, among other things, that the motions of the visible Brownian particles are caused by collisions with the molecules making up the dilute liquid in which those visible particles are suspended" (Achinstein, 2002, 472). With this equation he was able to calculate *Avogadro's number*, which he understood as the number of molecules in one gram-molecule of hydrogen (i.e., the quantity of hydrogen whose mass in grams is numerically equal to the mean molecular mass of hydrogen). Perrin used a number of different experimental arrangements to produce data yielding distinct but convergent calculations of Avogadro's number N ($\approx 6 \times 10^{23}$). Perrin concluded that

> Even if no other information were available as to the molecular magnitudes, *such constant results would justify the very suggestive hypotheses that have guided us* [including that molecules exist], and we should certainly accept as extremely probable the values obtained with such concordance for the masses of the molecules and atoms. … The objective reality of the molecules therefore becomes hard to deny. (Perrin, 1916, 105, quoted in Achinstein, 2002, 473; emphasis in original)

Achinstein's essay reconstructs the reasoning whereby Perrin arrives at this conclusion. He argues that Perrin's conclusion is indeed a realist one, in that he not only asserts the existence of unobservable entities (molecules), but he does so in opposition to empiricist scientists of his day, such as Pierre Duhem, Ernst Mach, Wilhelm Ostwald, and Henri Poincaré, who "raise general methodological objections to inferences from what is observed to what is unobservable" (Achinstein, 2002, 493). According to Achinstein, Perrin used a *causal eliminative* argument, in which one begins by listing all of the causes of some observed phenomenon E that are possible "given what is known" (alternatively, this is a list which, it is highly probable, includes the cause

of *E*). If, then, one can conclude of all but one of these possible causes that they do not, in fact, cause *E*, then one can justifably conclude that (probably) the remaining possible cause of *E* is its actual cause. Perrin, Achinstein argues, had good reason to believe that he could, through a combination of experimental and statistical arguments, rule out all other possible causes of Brownian motion sufficiently to claim that it was at least more probable than not that it was caused by the motion of molecules. His subsequent derivation of Avogadro's number then raised this probability even higher.[7]

Achinstein attempts to anticipate and respond to several anti-realist objections to his argument. One empiricist objection deserves particular attention. According to this argument, empiricism requires us not to draw inferences beyond what would be susceptible to either support or correction by possible future evidence. Claims about things that are unobservable cannot be supported or corrected by data concerning observable things because we cannot know that the conclusions we draw concerning the observable things will also hold for the unobservable things. Suppose that we are interested in sampling from things of type *A* and observing whether they have some property *B*. Even if we find that all of the members of our large and varied sample of *A*s have *B*, we cannot infer that *all* things of type *A*, including those that are unobservable, have *B*, since it may be that being observable is a *biasing condition* with respect to *B*. That is, perhaps *B* is a universal property among observable *A*s but is absent or rare among unobservable *A*s (assuming the latter even exist).

An argument like this would warrant denying empirical credentials to claims about unobservables: such claims require an extrapolation from data regarding observable things to conclusions about unobservables that is unwarranted in principle. But Achinstein notes that the principle required to make this argument is either self-defeating or arbitrary.

Suppose that the anti-realist argument from biased sampling is based on this principle: If all members of a sample of *A*s share some property *P* that is not shared by all *A*s outside of the sample, then no conclusion regarding all *A*s is warranted. That would suffice to underwrite the anti-realist conclusion

[7] Achinstein employs a species of Bayesian probability; a frequentist would reject such probability statements. His interpretation of probability departs from the usual Bayesian one, however. For him, the probability of a hypothesis is a measure, not of the degree of belief the agent does or should hold in the hypothesis, but the degree to which it is reasonable to believe the hypothesis. See his (2001) for an explanation and defense of this view and his general theory of evidence.

about observability (observability would be the property *P*), but it is so strong that *no* inference from a sample to a population would be warranted, since one can always find *some* such property for any sample (such as 'observed in Dr. Carlsson's lab' or 'located in the northern hemisphere').

If this very strong but self-defeating principle is not behind the anti-realist appeal to biased sampling, what principle is? While being observable may be a biasing condition for some inferences, the anti-realist argument here considered requires us to believe that it always is. That being observable *might* be a biasing condition should not by itself lead us to reject inferences concerning unobservables, since many other properties also might. Indeed, as Achinstein notes, we can go further and note how experimentalists deploy strategies for addressing worries about the warrant for extrapolating into the realm of the unobservable: Although one cannot observe the unobservable, "[o]ne can vary conditions or properties in virtue of which something is observable (or unobservable)" such as size, proximity, duration, and interactions with other bodies (Achinstein, 2002, 483). Perrin seems to have employed just such a strategy in his own experimental arguments. Speaking of the particles of gamboge he used, he wrote "*We shall thus be able to use the weight of this particle, which is measurable* [and which Perrin did vary by preparing several different emulsions of gamboge], *as an intermediary or connecting link between masses on our usual scale of magnitude and the masses of the molecules*" (Perrin, 1916, 93–94; emphasis in original).

10.6.3 Experiment and realism about positrons

Although Hacking's attempt to drive a wedge between realism about theoretical claims and realism about entities may not succeed, Massimi's and Achinstein's arguments support his insight that attention to experimental practice adds to the resources for vindicating a realist attitude in science in at least some instances. Moreover, even if experimental arguments for the existence of unobservable entities are not entirely theory free, they often allow one to draw conclusions about unobservable entities at a stage when the theoretical description of those entities remains unresolved by evidence (or perhaps no candidate for a theoretical description has even emerged).

Consider Carl Anderson's arguments for the existence of a positive electron. Anderson's findings could not serve as conclusive support for Dirac's theory. Even when Blackett and Occhialini developed more detailed evidence,

they denied that they could use their data to test Dirac's theory directly. But Anderson did provide an argument for the existence of *something* that, though unobservable, possessed particular attributes: it had a mass similar to that of an electron, and it had a charge opposite that of an electron. Anderson's argument drew upon some theoretical claims: particularly laws describing the effects of a magnetic field on a charged particle. He employed reasoning based upon previous experimental determinations such as "well established and universally accepted" descriptions of the range of protons of a given energy in a cloud chamber (C. Anderson, 1933, 491).

Indeed, drawing upon the previous chapter, we might construe Anderson's argumentation along the lines of an argument from severity in Mayo's sense: he considered what he took to be all of the alternatives to the conclusion that he drew that were possible given what he knew, and showed how under any alternative possibility, it would be very unlikely that one would observe an image that fit so well with the hypothesis of a positively charged electron-like particle (Staley, 1999a). Following Anderson's work, subsequent experimentation would both strengthen the argument for this fairly weak conclusion and provide a basis for stronger claims about the properties of positrons (Roqué, 1997).

We can certainly grant that in 1933 Dirac's theory gave 'the best explanation' of Anderson's observations, but did that provide good grounds for believing that theory? In accordance with Hacking, we can regard Anderson as having given an argument for the existence of positrons that we need not regard as an IBE. But Anderson's argument was also not based on being able to *use* positrons for some experimental purpose; it is not an engineering argument. By Hacking's criterion, a justified realist attitude toward positrons would have to wait for the advent of experiments using beams of positrons as probes of subatomic structure.[8]

10.7 Structural routes to realism

We have seen how different forms of realism arise from adopting a realist attitude toward different aspects of a theory. One version of this idea

[8] In the early 1960s an Italian collaboration built the first particle accelerator that used colliding beams of electrons and positrons, the AdA (*Annello di Accumulazione*) (Bernardini, 2004).

differentiates between the entities postulated by a theory and the structural relations between them. Beginning in the late 1980s, John Worrall began advocating *structural realism* (though he regards Henri Poincaré and Pierre Duhem as earlier advocates). Structural realists adopt a realist attitude toward the *form* or *structure* of certain theories, but not to the kinds of *things* the theory postulates. This theory began in its *epistemic* form, in which ontology is considered a matter about which one ought to remain agnostic. A more recent *ontic* version has recently emerged claiming that when it comes to fundamental ontology, *structure is all there is*.

10.7.1 "The best of both worlds": epistemic structural realism

Surveying the terrain of the long-standing dispute between realists and anti-realists, Worrall finds that both sides have both a compelling positive argument and a fatal flaw. Moreover, each position's strength reflects the other's weakness. On the realist side, Worrall acknowledges that the no-miracles argument exerts a strong intuitive pull. It *does* seem that some of scientists' empirical successes are inexplicable unless the theories responsible for those successes are getting something right. Prevailing anti-realist philosophies have no adequate alternative account for this great success. On the anti-realist side, it seems that even in fairly recent history the mature disciplines have undergone upheavals that have overthrown understandings of phenomena that had seemed settled and put in place of those understandings completely transformed interpretations of the natural world. It would seem hubristic to claim that all of that is behind us and that we will not be subjected to such revolutions in the future; rather we should *expect* such transformations to continue. No realist philosophy, even if we allow the invocation of 'approximate truth' seems to do justice to this historical pattern.

Worrall proposes that epistemic structural realism promises "the best of both worlds" by articulating a view that gives due consideration to the no-miracles argument, but can accommodate conceptual revolutions in science. Consider, for example, the problem of the nineteenth-century wave theory of light. This theory generated some remarkable successes. For example, Augustin Jean Fresnel gave it a mathematical formulation in an essay he submitted to the French Académie des Sciences for a prize competition in 1819. One member of the committee, Siméon Denis Poisson, used Fresnel's theory to calculate the surprising result that, given the right

configuration, one would find a bright spot at the center of a shadow cast by an opaque disc. The chairman of the committee, François Arago, decided to test this unexpected consequence of Fresnel's theory, and found the predicted bright spot. Fresnel won the prize. But Fresnel's theory represented the propagation of light as the movement of vibrations in an all-pervading mechanical aether that current physics denies. For the scientific realist, the wave theory of light poses a problem as an empirically successful theory using theoretical terms referring to things that do not exist. (We have already noted the 'divide and conquer' response to this problem in Section 10.5.2.)

But the history of the theory has an important feature for which a purely anti-realist attitude seems inadequate: When the wave theory based on a mechanical aether was supplanted by Maxwell's theory of electromagnetic fields, the mathematical equations theorists had developed to describe the undulations of the aether *remained*, but they now described the vibrations of electromagnetic fields. The progression of theories of light thus exhibits *preservation of structure* alongside a significant discontinuity in its description of entities. Worrall quotes a telling passage from Poincaré discussing the equations of Fresnel's theory:

> these equations express relations, and if the equations remain true, it is because the relations preserve their reality. They teach us now, as they did then, that there is such and such a relation between this thing and that; only the something which we then called *motion*, we now call *electric current*. But these are merely names of the images we substituted for the real objects which Nature will hide for ever from our eyes. The true relations between these real objects are the only reality we can attain ... (Poincaré, [1905]1952, quoted in Worrall, 1989b, 118)

In other cases, the mathematical structure of an earlier theory might be subsumed within that of a later theory as a *special case*. We find this pattern in the replacement of Newton's dynamics by the dynamics of Einstein's special relativity. Kuhn argues (as discussed in Chapter 5) that the change from Newton's physics to Einstein's physics is so great that even though the same terms appear in both theories, their meaning is changed. So Newton's physics as a whole (with its understanding of mass as a quantity independent of one's frame of reference, for example) is not really a special case of Einstein's physics. But even if Kuhn is right about this, the structural realist's more

modest claim remains valid: the mathematical form of Newton's laws can be derived from Einstein's laws as a special case where relative velocities are very small compared to the speed of light. This shows that in Einstein's theory, the relations between quantities that Newton's equations describe remain (within a certain domain) unchanged. Einstein's theory revises what those quantities signify, but Newton's theory gave the correct description, within a restricted domain, of how they relate to one another.

Dirac's hole theory, one more time

In one respect, the case of Dirac's theory and its successful prediction of the positron seems to lend support to structural realism. If one consults a current textbook on electrodynamics, one will find equations with apparently the same mathematical form as in Dirac's own presentation of his theory, in spite of the fact that Dirac regarded his theory as a description of the relations between particles that fill space even in a vacuum, whereas the current theory describes relations between fields.

Once again, however, the success of Dirac's peculiar theory challenges our attempts to craft a systematic philosophical understanding. As Pashby notes, interpreting the transition from Dirac's hole theory to our current understanding of quantum electrodynamics as a story of structural preservation tells only part of the story. Although we will not here delve into the mathematics required for a full account, the crucial point is that in spite of the formal similarity between Dirac's own version of his equation and the modern understanding of it the *mathematics* of the two equations are in fact quite different. In Dirac's theory, his equation describes the evolution of a *wave function* that determines the probability of finding a particle in a given location at a given time. But the modern equation applies instead to fields that assign *operators* to particular locations in spacetime. These operators then determine the probability that a particle will be *created* or *annihilated* at a given location. The difference here cannot be easily interpreted as simply a change from one kind of thing to another kind of thing while keeping the relations between the things the same. Pashby concludes: "theory replacement involves structural discontinuity as well as structural preservation, and since these discontinuities may reflect the problematic ontological or metaphysical shifts that the structural realist had hoped to avoid, she had better account for them too" (Pashby, 2012, 470).

10.7.2 Ontic structural realism

The structural realist position we have been discussing has come to be called *epistemic* structural realism to distinguish it from the more recently emerged *ontic* structural realism. Epistemic structural realism argues that we can safely regard our best theories as mapping an *aspect* of the world: its structure. But "the *nature* of the basic furniture of the universe" will remain forever elusive to us (Worrall, 1989b, 122).

Ontic structural realists, however, argue that the structural realist understanding of our best current theories in physics provides us with just such an understanding of the "basic furniture of the universe" (or, perhaps better, how the furniture has been arranged) and that what it tells us is that it is precisely *structure* that is fundamental to the makeup of the universe. In one version of this idea, at the most basic level of the makeup of the universe, there are no *things* that exist independently of the relations that hold between them. Rather, it is the relations (i.e., the structures) that are fundamental (Ladyman et al., 2007).

Epistemic structural realism attempts to carve out a realist understanding of the success of scientific theories that disentangles them from metaphysics. Ontic structural realism invokes that same understanding as a way of *doing* metaphysics.

10.8 Conclusion

Scientists have developed remarkably sophisticated methods and technologies for learning about the world. It is perhaps shocking, therefore, that there is so little agreement about *what* we learn from their expert use of these methods and technologies. Imagine that you went to a restaurant run by a world-famous chef with a state-of-the-art kitchen, serving a spectacular menu of delicious selections, but none of the staff could agree whether any of the items on that menu would actually nourish you! On the other hand, the extended conflict has encouraged both sides to refine their arguments, and some untenable positions have been abandoned.

These disputes have generated a multitude of contrasting arguments and positions, such that one cannot hope to do justice to every important contribution in a single chapter of a book such as this. Let me conclude by just mentioning a few authors whose writings deserve further attention.

Ernan McMullin argues for a nuanced form of realism that pays special attention to the difference between physics (where perhaps the case for anti-realism is strongest) and other sciences like geology. McMullin's argument emphasizes the ways in which science relies on metaphors (McMullin, 1984).

In order to improve on Laudan's pessimistic meta-induction *Kyle Stanford* advocates a historical argument for a kind of cognitive limitation leading to a recurring underdetermination problem. According to this argument, it is not simply the occurrence of empirically successful theories that are not approximately true, but the many historical examples of scientists failing to conceive of *relevant alternatives* to their favored explanatory theories that should persuade us to be anti-realists (Stanford, 2006).

Arthur Fine seeks to transcend the dichotomy between realism and anti-realism with what he calls the "Natural Ontological Attitude" (NOA). He considers both realism and anti-realism to be "inflationary" philosophical adjuncts to science that seek to tie the legitimation of science to either metaphysics (realism) or epistemology (anti-realism). By contrast NOA includes a "pro-attitude" towards science and even towards science yielding belief in the truth of claims about unobservables, but it avoids theorizing about truth itself or about the aims of science, which it regards as a "chimera" and an artifact of the philosophical dispute to which we have devoted our discussion in this chapter (Fine, 1984, 1986).

Howard Stein follows a rather different path from Fine's toward the idea that scientists in practice incorporate aspects of both realism and instrumentalism in a way that renders them compatible (Stein, 1989).

11 Explanation

11.1 Introduction

That we use scientific theories to *explain* things is a matter of broad (though incomplete) agreement.[1] But what is it to explain something? Intuitively, explanations enable us to *understand* the phenomena we observe, where understanding involves something more than merely knowing *that* something occurs. Intuitively, again, that *something more* seems to involve a knowing *why* or knowing *how* the phenomenon occurs. But none of these intuitive ideas helps very much with an analysis of what constitutes an explanation (or what constitutes a good explanation). For what constitutes understanding? and what does it take for an explanation to help us achieve it?

This is a very old question in philosophy. Aristotle insisted that in order to "grasp the 'why' of something" one should understand each of its four different kinds of *cause*: what it is made of, its form, its source of change or stability, and its end or purpose.

Aristotle's four causes are really four kinds of explanation, and they reflect his own approach to explaining natural phenomena. The discussion of explanation within contemporary philosophy of science likewise draws motivation from the explanatory practices of contemporary science; it really begins with Carl Hempel and Paul Oppenheim's *Deductive-Nomological* (D-N) model of explanation, which constitutes the core of the *Covering Law* model. We will therefore begin with the Covering Law account and objections to it before turning to several prominent alternative accounts.

[1] Pierre Duhem is among the dissenters. He rejected the view that the aim of physical theory is to explain phenomena, holding that to explain "is to strip reality of the appearances covering it like a veil, in order to see the bare reality itself." This task belongs to metaphysics (Duhem, 1991, 7).

11.2 The Covering Law model

We can start with a simple example:

> A mercury thermometer is rapidly immersed in hot water; there occurs a temporary drop of the mercury column, which is then followed by a swift rise. How is this phenomenon to be explained? The increase in temperature affects at first only the glass tube of the thermometer; it expands and thus provides a larger space for the mercury inside, whose surface therefore drops. As soon as by heat conduction the rise in temperature reaches the mercury, however, the latter expands, and as its coefficient of expansion is considerably larger than that of glass, a rise of the mercury level results. (Hempel and Oppenheim, 1948, 135–136)

In every explanation we can distinguish between the *explanandum*, a statement describing what is to be explained, and the *explanans*, the statements that tell us what explains the explanandum. In this example the explanandum has two parts: First, that the height of the column of mercury in the thermometer is briefly lowered; second, that the height of the mercury column subsequently increases.

The explanation quoted above leaves out all mathematical details. It does not tell us how much the column of mercury falls or rises, or how long it takes to do these things. The laws that describe the processes of heat conduction and thermal expansion are invoked implicitly but not stated in their full quantitative form. But Hempel and Oppenheim nonetheless think that the implicit references to these more detailed and quantitative expressions point us to the requirements of a "sound" explanation, which they divide into *three logical requirements* and *one empirical requirement*:

Logical requirements:

DNL1 The explanans must deductively entail the explanandum. A D-N explanation is a deductive *argument* in which the explanans constitutes the premises and the explanandum constitutes the conclusion.

DNL2 The sentences that make up the explanans must include some sentences expressing general *laws*. (Typically, though this is not required, they will also include some sentences not expressing laws but describing so-called *initial conditions* that in conjunction with the laws make possible the derivation of the explanandum.)

DNL3 The explanans must have *empirical content*, meaning that "it must be capable, at least in principle, of test by experiment or observation" (Hempel and Oppenheim, 1948, 137).

Empirical requirement:

DNE1 The sentences constituting the explanans must be true.

Hempel and Oppenheim summarize these requirements through a general schema for D-N explanations:

$C_1, C_2, \ldots, C_n$	(statements of antecedent conditions)
$L_1, L_2, \ldots, L_m$	(general laws)
E	(description of the phenomenon to be explained)

Here the explanans are given above the line, the explanandum below, and the former must deductively entail the latter.

According to the D-N model, an explanation is a certain kind of deductive *argument*. Hempel also argued that every explanation could function equally well as a *prediction* for someone who knows the explanans prior to the explanandum. Explanation and prediction share the same logical structure, and they do so because that logical structure enables explanations to satisfy the criterion of *nomic expectability*: an explanation must show that the phenomenon described by the explanandum *was to be expected* in virtue of the laws applicable to the situation in which that phenomenon occurred.

Hempel developed another model for explanations drawing upon laws that are statistical in nature, which he called the Inductive-Statistical (I-S) model. Here it will suffice to note that it resembles the D-N model, but with the universal laws of D-N replaced with *statistical laws*, from which (along with the statement of antecedent conditions) one can derive, not that the explanandum must be true, but that the explanandum has a very high probability, given the explanans. In this way, the I-S model also meets the nomic expectability requirement, though less perfectly. I will use the term *Covering Law* (CL) model to refer to the combination of the D-N and I-S accounts.

As we noted, our thermometer example does not exactly match this schema as stated. It constitutes instead an *explanation sketch*, meaning that it does not state explicitly everything that an explanation must include, but gives enough of an outline that we can see what would need to be filled in to make a complete explanation.

The idea of an explanation sketch helps the CL model to accommodate our frequent use of explanations that do not explicitly include all of the elements that the model requires. Especially in the social sciences, in history, and in our attempts to understand the purposive actions of human beings, we do not often find statements of general laws from which the description of the

phenomena we seek to understand can be deduced, even with the help of additional background information or initial conditions. But Hempel and Oppenheim nonetheless contend that such explanations are really explanation sketches that, although incomplete, make "reference" to both specific antecedent conditions and general laws (Hempel and Oppenheim, 1948, 141).

11.2.1 An example: the Hodgkin and Huxley model of the action potential

To further explore the motivations and the limitations of the CL model, let us consider a somewhat extended example from the history of neuroscience.

Most cells in an organism can be thought of as batteries. That is, there is a voltage, or *electrical potential difference* (V_m) between the inside surface and the outside surface of the cell membrane, due to differences in the concentrations of ions (electrically charged atoms or molecules) on either side. In certain kinds of cells, that voltage sometimes changes significantly over very brief intervals following a characteristic pattern. Such changes are called *action potentials*, and they are due to changes in the permeability of the membrane that allow ions to diffuse across it. The movement of ions through the membrane constitutes an electrical current. Although action potentials can be found in a variety of kinds of cells, our present example concerns their role in neurons.

At the beginning of an action potential, the neuron is in a resting state in which the potential difference between the inside surface and the outside surface of the membrane is -60 mV to -70 mV (about 1/20 of the voltage of an AA battery). This is the *resting potential*, V_{rest}. Should V_m rise above a certain *threshold* value (~ -55 mV), it will then rapidly rise to a peak of +35 mV to +40 mV, and then rapidly decline to a value a little below the resting potential, thus initiating the *refractory period* during which the cell is less excitable. Gradually, V_m will then rise back to the level of V_{rest} (Figure 11.1).

In neurons, the electrical current produced in an action potential is chiefly due to the flow of sodium (Na^+) and potassium (K^+) ions across the membrane. Other ions contribute smaller amounts to the total current, and the total of these smaller contributions is called the *leakage* current. For each of the ions Na^+ and K^+, as well as for the leakage current, there is a distinct value of the potential difference at which that ion is in equilibrium, i.e., there is no net current in either direction due to the movement of that ion.

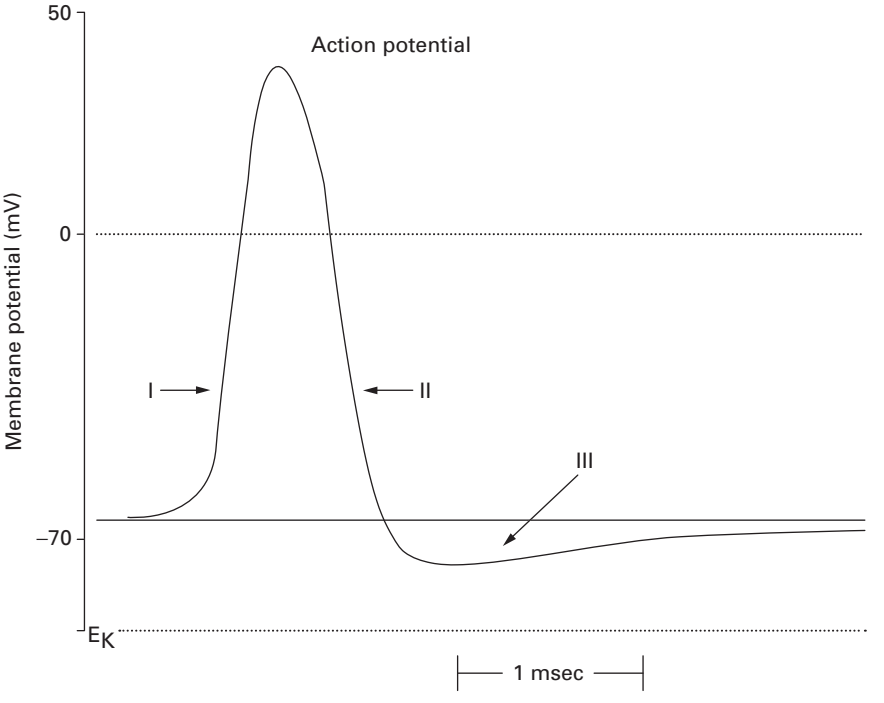


Figure 11.1 A schematic drawing of an action potential indicating the main features of the changing voltage across the cell membrane: (I) a rapid rise of V_m, followed by (II) a steep decline to values below that of the resting potential and (III) an extended refractory period. (Craver 2007, 50; by permission of Oxford University Press)

Alan Hodgkin and Andrew Huxley (henceforth HH) developed a mathematical model of the action potential in neurons based on a series of experiments on an unusually large nerve fiber, or *axon*, found in squids. HH published the results of these experiments in a series of papers in 1952 in the *Journal of Physiology*, summarized in (Hodgkin & Huxley, 1952). The mathematical model they developed employs a *total current equation* that sets the total current equal to the sum of four contributing terms:[2]

$$I = \underbrace{C_M dV/dt}_{\text{capacitance current}} + \underbrace{G_K n^4 (V - V_K)}_{\text{potassium current}} + \underbrace{G_{Na} m^3 h (V - V_{Na})}_{\text{sodium current}} + \underbrace{G_l (V - V_l)}_{\text{leakage current}} . \quad (11.1)$$

[2] I follow Craver (2007, 51) in writing the equation as Hodgkin and Huxley wrote it, rather than as it will be found in contemporary textbooks.

In this equation, V is the difference between V_m and V_{rest}. G_K, G_{Na}, and G_l are the maximum conductance (the inverse of electrical resistance) values for the three ionic currents. V_K, V_{Na}, and V_l are the respective differences between the equilibrium potentials for the three ionic currents and V_m. The first term, the capacitance current, depends on the capacitance of the membrane, C_M, which measures the ability of the membrane to store oppositely charged ions on opposing sides, and on the rate of change of the voltage, dV/dt. The coefficients n, m, and h depend on V and vary with time, and HH give a series of first-order differential equations for them (Hodgkin & Huxley, 1952, 518–519). Those additional equations are crucial to the application of the HH model, and references to the 'HH model' should be understood to include them in addition to Equation (11.1).

Equation (11.1) looks like a law, stating that whenever the quantities on the right-hand side of the equation take on certain values, the quantity on the left-hand side – the total current across the membrane – will take on a value determined by the equation itself. Moreover, HH appear to use it to explain certain phenomena regarding action potentials. They summarize their paper by stating that Equation (11.1), along with the associated differential equations for n, m, and h, "were used to predict the quantitative behavior of a model nerve under a variety of conditions which corresponded to those in actual experiments." They then enumerate eight properties of the action potential for which "good agreement was obtained," such as the "form, amplitude and threshold of an action potential under zero membrane current at two temperatures," the "form, amplitude, and velocity of a propagated action potential," and the "total inward movement of sodium ions and the total outward movement of potassium ions associated with an impulse" (Hodgkin & Huxley, 1952, 543–544).

According to Hempel, only the pragmatic distinction between deriving a statement prior to knowing it to be true and after knowing it distinguishes prediction from explanation. So if we view HH's summary through the lenses of the CL model, we will understand the claim that they used Equation (11.1) and its associated equations to *predict* these aspects of the action potential to entail that we can likewise use those equations to *explain* those phenomena. Craver writes, "There is perhaps no better example of covering-law explanation in all of biology" than the HH model (Craver, 2008, 1022). Later in this chapter, we will consider reasons why we might, as Craver himself advocates, wish to revise this judgment.

11.2.2 Objections to the CL model

Objections to the CL model have circulated long enough to form a set of criteria used to judge other models of explanation. Here we present four.

Explanations without laws

The first objection holds that the requirements of the CL model are not *necessary* for explanation. We have already noted that many examples of explanations even within science do not explicitly state laws and initial conditions that suffice for the deduction of the explanandum. The CL model holds that we should expect such efforts to *implicitly* refer to a more detailed explanation that does include such statements, in the way that a sketch of a figure shows us where a more detailed rendering would provide more information about the appearance of the figure sketched. Can we explain something without either explicit or implicit reference to a general law?

Michael Scriven gives the example of a person who, while reaching for a cigarette, knocks over an ink bottle, causing the ink to spill onto the carpet. The person can then give the following explanation for how the stain appeared on the carpet: she knocked the ink bottle over. "This is the explanation of the state of affairs in question, and there is no nonsense about it being in doubt because you cannot quote the laws that are involved, Newton's and all the others; in fact, it appears one cannot here quote any unambiguous true general statements, such as would meet the requirements of the deductive model" (Scriven, 1962, 198).

The ink bottle example reveals that we not only give and accept explanations that do not explicitly cite laws, we do so when we could not state any "unambiguous true general statements" (let alone *laws*) that would enable us to deduce a description of the phenomenon to be explained (thus ruling out that we are somehow *implicitly* appealing to such a law). For example, the generalization 'Whenever an ink bottle is bumped, an ink stain appears on the carpet' might suffice for the required derivation, but is clearly not true. We might try to make the statement more specific: 'Whenever an open ink bottle is bumped hard and falls over, and there is ink in the bottle, and the bottle sits on a table, below which is a carpet, then an ink stain appears on the carpet.' But this is still not true, since one might intervene to set the ink bottle aright before enough ink spills to run off the table, or one might

quickly cover the carpet with a towel, or the dog might lie down in the path of the spilling ink, or … It is not obvious that there is any non-trivial way to build all of this into the needed generalization, and the more that we do build these conditions into the statement of the generalization, the less the resulting statement looks like a statement of a *general law*. (Indeed, just what does constitute a law is a matter of ongoing debate to which we could easily devote an entire chapter.[3])

Nomic expectability

Another important criticism of the CL model strikes directly at one of its chief motivations: nomic expectability. Some seemingly satisfactory explanations do not show that the phenomenon to be explained was to be expected in light of the applicable laws, but only cite circumstances that are *causally relevant* to its occurrence. Consider another example from Michael Scriven (1959): Some people with syphilis develop a form of paralytic dementia known as *paresis*. Most people who contract syphilis do not develop paresis, but in those who do, it is their having contracted syphilis that explains why they developed paresis. At any rate, this is the best explanation we can give. Since having contracted syphilis explains, but does not render predictable, that a person develops paresis, nomic expectability is not a necessary condition for an explanation.

Advocates of a CL approach might object that of course there must be some difference between syphilis patients who develop paresis and those who do not, and that a full explanation of an individual case of paresis would cite those additional distinguishing factors. We simply do not yet know enough to give that explanation.

The problem with this response, however, is that the CL model is committed not just to saying that it would be a *better explanation* if we could show that we should expect paresis in that individual, but that citing the patient's prior condition of having syphilis does not explain his paresis *at all*. And in many cases – indeed in many fields of science – it simply is not obvious that we should expect to be able someday to give explanations that meet the criterion

[3] See (Armstrong, 1983), (Cartwright, 1983), and (D. Lewis, 1986) for some influential discussions. See (Lange, 2009) and (Roberts, 2008) for some recent proposals that have garnered attention.

of nomic expectability; at any rate, whether we can *now* explain something ought to be independent of whether we will in the future be able to explain it better. Scriven gives his example in the course of arguing that Darwinian evolutionary explanations of survival by appealing to fitness exemplify the dispensability of the nomic expectability requirement.

Asymmetry

Many scientific laws express some kind of *equivalence* between quantities used to characterize a physical system. Such laws can be applied equally in either direction. For example, players of wind instruments create sound by vibrating, either with their lips (brass instruments) or with a reed (woodwinds), the air in a metal tube. At resonant frequencies, this creates a standing wave in the horn. We can use a simple expression to approximate the frequency f of the standing waves in a resonant metal cylindrical pipe that is closed at one end and has length L: $f = (nv)/(4L)$, where n is an odd number and v is the speed of sound in air. A D-N explanation of the frequency of the sound waves coming from a given resonating horn would cite this expression as well as a description of the length of the horn.

The *asymmetry* problem for the CL account arises because we can run such derivations in reverse (Bromberger, 1966): a derivation of the length of the horn from the sound waves it produces when resonating satisfies the requirements of the CL model of explanation just as well. But it seems that explanatory relationships do not share this symmetry. We appeal to the length of the pipe to answer the question 'why does it produce these particular sound waves when resonating?' (Call this the 'length-to-notes' explanation.) But we do not, in the same way, appeal to the sound waves produced to answer 'why does it have the length that it does?' (the 'notes-to-length' explanation).

I say "in the same way" because we might well give a certain *kind* of notes-to-length explanation for the case of horns or other musical instruments using tubes. In those cases, the tube was given the length that it has by the manufacturer so that it would produce the particular kind of sound waves that it does. But here the explanation invokes both the properties of the sound waves and the intention of the manufacturer to produce an object with certain tonal properties, whereas no such considerations need enter into the length-to-notes explanation. The explanatory relationships are not symmetrical in the same way as the derivational relationships. Moreover the

length-to-notes explanation applies to all tubes, but the notes-to-length explanation does not apply to natural or accidental resonant cavities (like conch shells or moonshine jugs).

Irrelevance

The final problem, like the asymmetry problem, challenges the *sufficiency* of the CL conditions for explanation. For any lawlike generalization that we might use in an explanation, such as 'every sample of sodium chloride (salt) crystals dissolves in water' we can add an irrelevant condition to the antecedent, such as being hexed by a magician. This yields a new, seemingly lawlike generalization: 'every hexed sample of sodium chloride crystals dissolves in water.' We can use this new generalization, just like the original one, to derive an explanandum statement describing the dissolution of a particular sample of hexed sodium chloride in water, but the derivation that appeals to the law about hexed salt does not really seem to explain why the salt dissolved (Kyburg, 1965).

The problem of irrelevance also arises in cases allowing the derivation of explananda by appeal to generalizations that are 'accidentally' true. Suppose that every member of a six-person band happens to read ancient Greek. This does not explain why a particular member of that band (Scott) reads ancient Greek, although we can derive the statement 'Scott reads ancient Greek' from this generalization along with the statement that Scott is a member of that band.

Of course, if we had settled criteria for deciding whether a given generalization is a law, we might be able to use those criteria to exclude these cases on the grounds that 'every hexed sample of sodium chloride dissolves in water' and 'every band member reads ancient Greek' are not really laws. But Hempel himself acknowledged that the determination of lawfulness remained something of an unsettled issue for his account.

11.2.3 The Hodgkin–Huxley model as CL explanation?

In discussing the Hodgkin–Huxley model of the action potential, we mentioned that it seems like a good candidate for a covering law explanation of the features of the action potential that HH discuss. After all, the equation appears to describe quantitatively the way that the total current across the membrane depends on the quantities on the right side of Equation (11.1),

and by extension on the quantities in the additional equations for the conductance factors in Equation (11.1). Moreover, HH state that they were able to use these equations to *predict* those features of the action potential in a way that yielded "good agreement" with the results of their experiments.

In a careful study of this episode, however, Carl Craver points out that, although their mathematical model can do *some* explanatory work, HH themselves denied that every part of that model that could serve a *predictive* function was thereby explanatory: "The agreement [between model and data] must not be taken as evidence that our equations are anything more than an *empirical description of the time-course* of the changes in permeability to sodium and potassium. *An equally satisfactory description of the voltage clamp data could no doubt have been achieved with equations of very different form, which would probably have been equally successful in predicting the electrical behaviour of the membrane*." Consequently, "*the success of the equations is no evidence in favour of the mechanism of permeability change that we tentatively had in mind when formulating them*" (Hodgkin & Huxley, 1952, 541; quoted with emphasis added in Craver, 2007, 53–54).

More specifically, the HH model does, on Craver's account, explain the current involved in the action potential in terms of voltage and conductance changes. But in Equation (11.1) the conductances for potassium and sodium ions are given by the products of the maximum conductances and some function of the time-dependent quantities n, m, and h. It is the additional partial differential equations that HH give for the evolution of these latter quantities over time that they say give only an "empirical description of the time-course" of these sodium and potassium conductances. They predict the outcomes of measurements, but do not give us a *causal* story about how those quantities change.

Craver's point is that, although HH's conductance equations seem to meet the requirements of the CL model, they do not explain the changes in conductance. Of course a defender of the CL model might simply insist that if those equations do not explain conductance changes, then they must not express laws.[4] Drawing upon the statements of Hodgkin and Huxley and of later

[4] Marcel Weber argues that the explanations that make use of the HH model derive their explanatory force not from the HH model itself but from the physical and chemical laws (describing the forces applied to charged particles, etc.) that underlie the HH model (Weber, 2005).

neuroscientists who have continued to study mechanisms of ion transfer between neurons, Craver imposes a causal requirement on explanations and concludes that what prevents the HH model from explaining the phenomena regarding action potentials is not that it does not express a law, but that, at least for the conductance changes, it does not show how the phenomena are "situated within the causal structure of the world" (Craver, 2007, 49).

Perhaps, then, we should consider another example in which we can see a phenomenon situated in the causal structure that brings it about.

11.2.4 Another example: curare poisoning

Early nineteenth-century British explorers in South America reported how natives hunted using arrows dipped in a poisonous plant residue called *wouralia*.[5] The natives they encountered in Guiana and Brazil knew that this deadly substance would not only quickly kill the monkeys that they hunted, but would bring certain death to the unfortunate hunter struck by an arrow gone astray.

There was no question that wouralia, which came to be known as *curare*, caused death, but how it did so posed a mystery. As lethal as curare was on the tip of an arrow, natives suffered no ill effects from eating curare-poisoned monkeys. Research published in the 1850s by physiologist Claude Bernard revealed that curare produces death only when it enters the bloodstream. Bernard determined that the structure of curare prevents it from being absorbed efficiently into the bloodstream through the stomach. Moreover, he established that the hearts of frogs poisoned with curare continued to beat for some time after they stopped breathing. Further research revealed that curare stopped one from breathing by somehow breaking the *connection* between motor neurons and the muscles responsible for breathing. (Curare left unimpaired the functioning of both the motor neuron itself and the muscle.)

Bernard was able to take us beyond knowing that curare causes death to knowing that it does so by interfering with the connection between motor neurons and muscles. But how does it do that? Modern neuroscientists have worked out the answer. Roughly, the explanation goes like this: Ordinarily, electrical stimulation of the motor neuron triggers the release of the neurotransmitter acetylcholine (ACh). When ACh released by the motor neuron

[5] See (Craver & Darden, 2013, ch. 1) for a very readable discussion of this episode. My own discussion follows theirs closely.

binds to ACh receptors in the muscle membrane, they open a pore through which charged ions can then pass, and it is this electrical signal – an ion current – that triggers muscular contraction. The active ingredient in curare (tubocurarine chloride) mimics the shape of ACh, binding to the ACh receptors without opening pores through which an ion current can pass. Motor neurons can no longer activate muscle contraction because the curare blocks their activation signals from being received by the muscle.

Craver and Darden summarize the story nicely:

> How does curare kill its victims? It enters the bloodstream and makes its way to the neuromuscular junction. There, it blocks chemical transmission from motor neurons, effectively paralyzing the victim. When the diaphragm is paralyzed, the victim cannot breathe, and animals that cannot breathe do not get the oxygen required to maintain basic biological functions. Thus is hunter's lore about the irrevocable effects of curare transformed into scientific knowledge of mechanisms. (Craver & Darden, 2013, 2)

Later in this chapter we will consider perspectives on explanation that emphasize *causation* and the *mechanisms* that produce phenomena.

11.3 Achinstein's pragmatic theory

Thus far we have rather carelessly failed to distinguish between the *act* of explaining, the *explanation that is given* in that act, and the *evaluation* of that explanation. Peter Achinstein insists that one must begin an account of explanation with an analysis of the act of explaining as fundamental, and that "the concept of an explanation (as product), and that of a good explanation, must be understood, in important ways, by reference to" the concept of an explaining act (Achinstein, 1983, 4).

11.3.1 Illocutionary acts

To explain, Achinstein tells us, is to perform an *illocutionary act*. Such acts are "typically performed by uttering words in certain contexts with appropriate intentions" (Achinstein, 1983, 16).[6] Other examples include promising, thanking, requesting, warning, and betting.

[6] The term comes from the philosopher J. L. Austin (1962). For more on illocutionary acts, illocutionary force and *speech acts* generally, see (Searle, 1969) and (Searle & Vanderken, 1985).

Consider a sentence of the form '*S* explains *q* by uttering *u*,' where *S* is a person, *q* expresses an indirect question (such as 'why the mercury rose in the thermometer'), and *u* is a sentence. Achinstein proposes three conditions, individually necessary and jointly sufficient, on the truth of a sentence of this form:

1. *S* utters *u* with the intention that his utterance of *u* render *q* understandable.
2. *S* believes that *u* expresses a proposition that is a correct answer to *Q*, where *Q* is the direct question whose indirect form is *q* (e.g., *Q* might be 'why did the mercury rise in the thermometer?').
3. *S* utters *u* with the intention that his utterance of *u* render *q* understandable by producing the knowledge, of the proposition expressed by *u*, that it is a correct answer to *Q*.

So, for example, consider the indirect question *c*: 'why there is an ink stain on the carpet,' the direct form of which is *C*: 'why is there an ink stain on the carpet?' Uttering the sentence *IB*: 'There is an ink stain on the carpet because I knocked over the ink bottle while reaching for a cigarette' might constitute an act whereby I explain why there is an ink stain on the carpet, provided that these three conditions are met. I must utter *IB* with the intention that my doing so will make understandable why there is an ink stain on the carpet. If I address my utterance to the dog simply to see him wag his tail, which I know he does regardless of what I say to him, then I am not explaining *c*; I am teasing the dog. Achinstein's analysis also requires that I must believe that *IB* expresses a correct answer to *C*. If I believe that the ink stain was actually produced by the dog knocking over the ink bottle, but I am protecting the dog from getting a scolding, then I am not explaining why there is an ink stain on the carpet; I am covering up why there is an ink stain on the carpet. Finally, my intention in uttering *IB* must be not merely to render *c* understandable, but to accomplish this by producing the knowledge that the proposition expressed by *IB* is a correct answer to *C*. This condition will not be met if, say, I know that although *IB* expresses a correct answer to *C*, you tend not to believe anything I say, but I also know that you compulsively look up scientific information on any topic that comes up in our conversation, and I intend, by uttering *IB*, to prompt you to go online to study the chemical processes whereby fabric becomes stained by ink.

These conditions thus allow Achinstein to avoid the *illocutionary force problem*. The utterance of any given sentence or group of sentences can be the

product of quite distinct illocutionary acts, depending on the context and intentions of the speaker. 'I will pay for everyone's dinner' might be either a promise or a prediction. Achinstein regards this as a problem for any theory of explanation that specifies an ontology of explanations as certain kinds of sentences or arguments independently of the *act* of explaining.

We noted at the beginning of this chapter that the notion of understanding seems closely related to that of explanation. One might worry, then, that Achinstein's first requirement, that S must intend to make q understandable by uttering u, might make his analysis either circular (if our account of understanding invokes explanation) or unclear (if we leave understanding undefined). He avoids this charge by offering an analysis of understanding that does not in turn depend on the concept of explaining or explanation. I will forego detailing here that analysis, other than to note that it endorses the intuition that understanding consists in more than simply knowing that something happened or happens. On Achinstein's account, understanding q is a certain *kind* of knowledge state with regard to q.

11.3.2 The ordered pair view

Calling Achinstein's account a *pragmatic* theory of explanation emphasizes the fact that for him explanation must be understood primarily as something that people (including scientists) *do*, with certain intentions and in certain contexts. For reasons such as the illocutionary force problem, Achinstein argues that one cannot analyze what it is to be an explanation except by reference to an explaining act.

Achinstein gives an *ordered pair* view of explanations in which candidate explanations consist of a pair $(x; y)$, where x is a certain kind of proposition and y is an action. Achinstein's conditions on $(x; y)$ being an explanation of q (an indirect question) given by S (a person) include that (a) y must be the act of explaining q, (b) that Q (the direct question of which q is the indirect form) must be a *content-question*, and (c) that x is a *complete content-giving proposition* with respect to Q. Rather than attempting to encapsulate the technical definitions of the terms used in conditions (b) and (c), let us hope that some examples will here suffice. (See Achinstein, 1983, 23–53 for a more detailed account.)

'Cause,' 'effect,' 'excuse,' 'explanation,' 'function,' 'meaning,' 'purpose,' and 'reason' are all examples of *content nouns*. Abstract nouns earn this title

in virtue of the role they play in *content-giving sentences* that express what their content is. Here are some examples of such sentences:

- The *cause* of the victim's death is that *he received a lethal dose of curare*.
- The *explanation* of the stain on the carpet is that *I was reaching for a cigarette and knocked over the ink bottle*.
- The *function* of the action potential is that *it makes communication between cells possible*.
- The *reason* you should turn off your phone in the concert hall is that *doing so will prevent people from considering you obnoxious*.

In each of these sentences a 'that …' expression *gives the content* of its content-noun. By contrast, the following sentences tell us something about their respective content nouns, but do not give the content of those nouns:

- The cause of the victim's death was discovered by performing an autopsy.
- The explanation of the stain on the carpet is not very scientific.
- The function of the action potential was understood before the mechanism behind it was.
- The reason you should turn off your phone in the concert hall is apparently something that you have failed to appreciate.

The sentences in the first list, but not those in the second, can be thought of as (possibly correct) answers to their respective *content questions*: What is the cause of the victim's death? What is the explanation of the stain on the carpet? etc.

Returning to Achinstein's conditions, then, an explanation of indirect question q by person S is an ordered pair of a proposition p and an act of explaining.[7] Requiring that the direct form Q of q must be a content question means that it must be a question asking for the content of a content noun. Requiring that the proposition p be a complete content-giving proposition with respect to Q means that S must use a sentence in the act of explaining that expresses a proposition giving the content of that noun ('completeness' involves additional technical requirements; Achinstein, 1983, p. 40).

[7] Achinstein also offers an alternative to the ordered pair view, the *no-product* view, that dispenses with the idea that acts of explaining yield products that are distinct from those acts (Achinstein, 1983, 98–100).

Let us pause for a moment to compare this account with the CL model. The most obvious point of difference is that Achinstein's account does not require, whereas the CL model does, that an explanation must include a law. Achinstein's model is simply less demanding than the CL account in this respect. Of course, one would need to cite a law in order to give an explanation as an answer to a content question that asks for one, such as 'What are the laws from which it follows that the voltage V in the final stage of an action potential will be lower than the resting potential?' Achinstein's account can thus handle cases like Scriven's ink stain.

In another respect, Achinstein's illocutionary account is *more* demanding. On the CL account, if an argument (a set of premises plus a conclusion) meets requirements *DNL*1–3 and *DNE*1, then it is an explanation, regardless of the context and intentions of anyone who might present it. But this is clearly not sufficient on Achinstein's account, since someone might utter such a set of sentences not as an act of explanation, but as part of some other illocutionary act, such as reminding or lamenting.

However, one might object that the CL model is not a model of the act of explaining but of *scientific explanation*. Conceding Achinstein's point that what counts as an act of explaining depends on the intentions and context of the action, one might insist that the CL model tells us what must be included in any utterance that is offered as an explanation of some explanandum q if the utterance is to be regarded as an adequately scientific explanation of q. This move Achinstein regards as a change of question. Instead of asking what is an explanation, we must now consider how to evaluate an explanation.

11.3.3 Evaluating explanations

Achinstein argues at length against the possibility of articulating universal rules for evaluating scientific explanations independently of the context in which the explanation is offered, aside from evaluating them as correct or incorrect. He gives the following analysis of the correctness requirement: "If (p; explaining q) is an explanation, then it is *correct* if and only if p is a correct answer to Q" (Achinstein, 1983, 106).

Beyond the requirement of correctness, what general instructions should explainers follow in order to give good scientific explanations? Achinstein articulates some conditions on the appropriateness of instructions that explainers might follow in the act of explaining q to a given audience, but

these conditions crucially involve facts about the audience in question, such as what the audience knows, what would help the audience to understand q, the kind of understanding the audience is interested in having of q, what would be valuable for them to understand about q, and what it is reasonable for the explainer to believe about all these things. In application, then, one could not determine appropriateness without some substantive empirical knowledge about the context in which one offers an explanation. In this way, evaluating explanations requires attention to context.

11.4 Unificationist theories

Many of the arguments against the CL have targeted its treatment of explanations as a certain kind of argument. But Philip Kitcher claims that behind their "official theory" of explanation logical empiricists such as Hempel held an unofficial view about scientific explanation: scientists seek to explain seemingly diverse phenomena by *unifying* them in a systematic way.

Moreover, according to Kitcher, one can reconcile this unification view of explanation with a rejection of the ontological thesis that explanations are arguments. Kitcher accepts instead Achinstein's ordered pair view of explanations (Kitcher, 1981, 509). Yet arguments remain important in Kitcher's account. Although an explanation is not simply a certain kind of argument, the act of giving a scientific explanation involves drawing upon arguments supplied by science. This leads Kitcher to a reformulation of the question about scientific explanation: "What features should a scientific argument have if it is to serve as the basis for an act of explanation?" (Kitcher, 1981, 510). Kitcher's broadly stated answer to this question is that it should be an instance of an argument pattern that scientists can use repeatedly to derive descriptions of a range of distinct phenomena, since "*Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again, and in demonstrating this, it teaches us how to reduce the number of types of facts we have to accept as ultimate (or brute)*" (Kitcher, 1989, 432; emphasis in original).[8]

[8] Kitcher echoes Michael Friedman's earlier account of explanation as unification (Friedman, 1974). But Kitcher seeks to repair technical problems that he pointed out in Friedman's account of unification (Kitcher, 1976).

Citing the fact that you knocked over the ink bottle while reaching for a cigarette might well explain why there is a stain on the carpet, but stating that you did so does not explain in a way that shows explicitly how the same kind of explanation could be given for stains from other kinds of liquids, or from other kinds of collisions.

Newton's theory of gravity and Darwin's theory of evolution by natural selection achieved tremendous unifications of phenomena. Showing that the same force, acting according to a single law, produces both the flow of ink downwards toward the carpet from the edge of the table and the motion of the moon around the earth contributed a new perspective that greatly increased our understanding of the motions of both terrestrial and celestial objects. Newton thus completed the dissolution of Aristotle's boundary dividing the cosmos into sublunar and superlunar realms containing distinct physical elements following different principles of motion. As Kitcher points out, Newton in fact proposed, and his successors pursued, an even more ambitious program of unification seeking to understand all physical phenomena in terms of similar laws describing other attractive or repulsive forces (Kitcher, 1981, 512–514).

In a different way, Darwin's theory, too, unifies a wide variety of biological phenomena. Although he does not give detailed explanations of characteristics of particular species in *On the Origin of species*, he does provide a pattern of argumentation that can in principle be instantiated with details for any given species. Instances of the pattern would cite facts about an ancestral population (its traits and their variability) and about the environment (determining what selective pressures were present), and would invoke laws of heredity and the principle of natural selection, thus showing how a particular trait would tend to emerge (Kitcher, 1981, 514–515; see M. Morrison, 2000 for a critical discussion of unification in both modern physics and Darwinian biology that challenges Kitcher's connection between unification and explanation).

11.4.1 Kitcher's theory

Kitcher provides a detailed account of just what constitutes an argument pattern, as well as a theory of explanatory unification that in principle would allow one to determine which set of arguments achieves the greatest unification at a given time, earning them a place among those arguments

that science provides for explanatory purposes, which Kitcher calls the *explanatory store*.

Roughly, Kitcher thinks of argument patterns as involving schematic sentences (think of these as the skeletons of sentences, showing their logical structure but without any flesh on their bones displaying what they are about). Such a *schematic sentence* will use placeholder terms like *X* instead of substantive terms like *curare*. Not just any way of filling in these placeholder terms will yield an instance of a given argument pattern. *Filling instructions* indicate how to go about replacing placeholder terms with substantive ones. In a Darwinian argument pattern, a placeholder might be replaceable by terms denoting selective pressures like predation or drought, but not by terms referring to divine intentions or vital forces. An argument pattern also needs to include a *classification* of sentences, telling which ones should serve as premises, which as conclusions, what inference rules one may use, and so on (Kitcher, 1981, 515–519).

Suppose that at a given point in a community's pursuit of inquiry they collectively accept a set of sentences *K*. Arguments allow one to derive some members of *K* from other members of *K*, and Kitcher refers to a set of arguments that does this for *K* as a *systematization of K*. The explanatory store relative to *K* is then the set of arguments that best systematizes or unifies *K*. Passing over the technical details, the gist of Kitcher's theory is that our assessment of how well a given set of arguments unifies *K* depends on the number of distinct argument patterns that are needed (fewer is better) to derive a number of accepted sentences (more is better). But these factors have to be evaluated in light of other relevant factors. Importantly, argument patterns must be *stringent*: the filling instructions must restrict the kinds of terms one may substitute for dummy terms in the pattern. Kitcher imposes this requirement to block cases of spurious unification by argument patterns allowing one to derive large numbers of sentences using trivial argument patterns, such as 'God wishes it to be true that *p* (where *p* could be any proposition whatsoever). Whatever God wants to be true is true. Therefore *p* is true.'

Let's revisit the asymmetry problem to see how the unificationist account works. Recall our example: the law relating the dimensions of a cylinder to the frequencies of the sound waves it produces in resonance allows one to derive the dimensions from the frequencies no less than the frequencies from the dimensions. Yet only the latter kind of derivation seems generally

explanatory. This, according to Kitcher, is because deriving dimensions from frequencies employs an argument pattern that is much less unifying than a competing approach, which he calls "origin and development derivations" (OD). These explain the dimensions of bodies by describing "the circumstances leading to the formation of the object in question" and "how it has since been modified" (Kitcher, 1981, 525).

The OD approach to explaining the dimensions of objects achieves greater unification than, say, deriving dimensions of an object from the acoustic effects of an object in resonance (let's call that RF for 'resonant frequency derivation') because we can apply it so much more widely. Even if an object might resonate, it often does not, and might never do so. Other objects have dimensions but no known law relating resonating frequencies to dimensions applies to them. In such cases, no appeal to an RF explanation is available, whereas an appeal to OD will be. Thus we may use OD to derive many more of our accepted sentences describing the dimensions of objects than we could using RF.

11.4.2 An objection to the unificationist account

Eric Barnes (1992) has objected to this treatment of the asymmetry problem by pointing to cases involving deterministic theories that describe the evolution of systems over time. *Deterministic* theories allow one, given a sufficiently precise description of the state of the system at a particular time, as well as of the forces acting on that system, to *predict* the future state of that system with a corresponding degree of accuracy, provided that the system described is *closed*.[9] We may call a system closed if our description of the forces acting on it includes all forces that will act on it over the relevant period of time that matter for the degree of accuracy we seek.

[9] To state a satisfactory definition of determinism is a non-trivial philosophical problem (Earman, 1986). For example, my definition appeals vaguely to a degree of accuracy of predictions 'corresponding' to a particular degree of precision in describing the initial state. This *correspondence* will vary from one theoretical system to the next. In *chaotic* systems, for example, very small degrees of imprecision in that initial state description can result in very large divergences amongst the possible evolutions of the state in the future. Yet chaotic systems are deterministic. See (Smith, 2007) for a lucid and accessible introduction.

Here is an example: We might apply Newton's celestial mechanics to the solar system to predict the positions and momenta of the planets at some time t_1, based on a description of the positions and momenta of the planets at an earlier time t_0 and the gravitational forces acting on them during the period $[t_0, t_1]$, assuming that no other forces will act on them during that time. We would thus use Newton's theory for a derivation using what Barnes calls a *Predictive Pattern*. Barnes's objection to unificationism notes that Newton's theory equally supports a *Retrodictive Pattern* allowing one to derive the state of the planets at t_0 from the same information about forces and the description of their positions and momenta at t_1.

There is no reason to think that the Predictive Pattern is more unifying than the Retrodictive Pattern. Both patterns use the same kind of information as inputs and yield the same kind of information as outputs, with only the temporal direction reversed. Yet that shift in temporal direction is associated with a shift in explanatory direction. The positions and momenta of the planets at a given time are *explained* by appeal to their positions and momenta at earlier times, not by their positions and momenta at later times. Since the two patterns unify equally well, the unificationist cannot account for this asymmetry.

11.5 Causal explanation

This temporal asymmetry maps very nicely, however, onto an asymmetry in *causal* relations: causes always precede their effects. Unificationists regard our practice of explaining effects in terms of causes as a byproduct of the unifying power of cause-to-effect derivations. An alternative view reverses the priority of these considerations, as when James Woodward states, "the causal order … is independent of and prior to our efforts to represent it in deductive schemes." Woodward cites Barnes's argument to support this claim (Woodward, 2003, 361). Although Woodward does not claim that all explanations or even all scientific explanations are causal, he does insist that providing such explanations is central to scientists' explanatory aspirations.

Clearly, such a view demands an account of causal explanation. Woodward provides an account of his own, which we will consider after first discussing one of its leading competitors.

11.5.1 Salmon's causal mechanical account

A common thread running through many accounts of causal explanation is that such explanations situate an explanandum in a network or nexus of causal relations.[10]

A particular version of this idea lies at the heart of Wesley Salmon's Causal Mechanical (CM) account of explanation (Salmon, 1984, 135–183). The central concepts in Salmon's theory are those of a causal process and of a causal interaction. *Causal processes* are physical processes capable of transmitting a mark in a continuous way, where a *mark* is a local modification of the structure of that physical process. Causal processes are extended in both space and time. A postcard traveling from St. Louis, Missouri to Millersville, Pennsylvania is a causal process. So is a baseball traveling from the pitcher's hand to the batter's bat, and then over the outfield wall to the windshield of a car in the parking lot. A shadow traveling along a wall is not a causal process. If I try to introduce a modification of the shadow by modifying it locally (i.e., doing something at the location of the shadow itself, rather than at the location of the object casting the shadow), such a modification will not be transmitted as the shadow moves from one location to another along the wall. The shadow exemplifies what Salmon calls a *pseudo-process*. A *causal interaction* is an intersection between two causal processes that modifies the structure of both of them, such as when I draw a pencil across the surface of the postcard, or when the baseball strikes the car windshield. A CM explanation of an event E then consists of a description of the causal processes and interactions that lead up to E (Salmon, 1984, 267–276; Woodward, 2003, 350–351). I explain the shattering of the car windshield by describing the ball's departure from the pitcher's hand, noting how the interaction with the bat modified its trajectory, and then describing how it interacted with the windshield.[11]

[10] The regularity with which discussions of causation invoke the language of networks, webs, weaves, tangles, and the like is striking and indicates how naturally we think of causal relations as a kind of *connection* between things. David Hume, nonetheless, insists that regular association alone relates causes to effects. For Hume, when we speak of something as a cause, we simply refer to "an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second" and "whose appearance always conveys the thought to that other" (Hume, [1748]1993, 51).

[11] Salmon later modified his view, replacing the CM model described here with a view of causal processes in terms of the transmission of conserved quantities (Salmon, 1994) that is a modified version of an account developed by Phil Dowe (1992, 1995).

The CM account seems well-suited for many explanations in physics[12] but, as Woodward has argued, many other kinds of causal explanations do not involve describing processes and interactions in the way that the CM account requires (Woodward, 2003, 354–356).

Consider, for example, the explanation offered in a paper published in the mid 1990s regarding the differences in unemployment levels between Spain and Portugal (Blanchard & Jimeno, 1995). Because of the significant and multifaceted similarities between the two countries, the fact that Spain had the highest unemployment rate in the European Union while Portugal had the second lowest rate constituted, the authors say, "the biggest empirical challenge facing theories of structural unemployment." The authors tentatively endorse the following explanation: "In Spain, high unemployment protection and unemployment benefits have led to small effects of labor-market conditions on wages. This led to large adverse effects of disinflation on unemployment in the first half of the 1980s. And high persistence since then explains why unemployment has remained high since." Lower unemployment benefits in Portugal have made wages more responsive to unemployment, resulting in "smaller adverse effects of disinflation on unemployment" and to less unemployment persistence (Blanchard & Jimeno, 1995, 216–217). Whether or not it is correct (see, for example, Ball, Mankiw, & Nordhaus, 1999 for another perspective), this explanation employs causal language, but without doing anything like what the CM model requires. The causes are policies regarding unemployment benefits. These policies make a difference to the decisions of individuals regarding the pursuit of employment, and the aggregation of many such individual decisions results in a particular pattern of economic activity. Although we might suppose that every interaction that contributes to these patterns involves physical human beings, the actions of whom can be traced down to the molecular level at which mark-transmitting causal processes and intersections between them can be identified, this explanation makes no reference to such processes. Moreover, we might not gain any additional insight into the explanandum phenomenon of disparate unemployment rates if we did take into account such processes.

[12] But see (Hitchcock, 1995) for a criticism that applies even to the use of CM in physics. Also, Robert Batterman (2001) explains how physicists often use *asymptotic explanation*, based on the analysis of what happens when some parameter in a theory approaches zero. Asymptotic explanations cannot be subsumed within the CM model.

11.5.2 Woodward's manipulationist account

If we want an account of explanation that applies not just to the processes described by physics but to causal relations operating at a variety of scales (of size, organization, and complexity), then we will need a different approach. James Woodward's account of causal explanation emphasizes a feature of causal relations that applies across, he argues, all disciplines: when one quantity has a causal influence on another, we can, in principle, make a difference to the latter by *manipulating* the former (Woodward, 2003).

Woodward's account of causal explanation is linked to his account of the meaning of causal claims. Rather than provide a complete discussion, the following brief remarks I hope will suffice to convey some idea of why Woodward's account has generated a significant discussion.

When we (rightly or wrongly) make a general causal claim such as 'smoking cigarettes causes heart disease,' 'taking birth control pills prevents pregnancy,' or 'listening to Mozart makes babies smarter,' what do we mean? For scientific purposes, we might sharpen the question: what meaning can we attribute to these statements that makes them amenable to empirical investigation through the collection of data?

One step toward answering this sharpened question would be to translate such claims into statements about the relationships obtaining between the values of *variables*. A reference to 'smoking' simpliciter conflates numerous dimensions of a complex behavior, different aspects of which can be represented by such different variables as number of cigarettes smoked per week, nicotine consumption per week, age at onset of smoking habit, duration of smoking habit, use or non-use of filters, brand of cigarette, and so on. (We can represent 'all or nothing' states – such as being pregnant or not – using *dichotomous* variables capable of taking either of two values, such as 0 for the absence of the condition or 1 for its presence.) Making one's claim of causal relevance specific to a particular variable can help to clarify more precisely the content of the causal claim (assuming the variable itself is well-defined). In Woodward's account of the meaning of causal claims, the causal relation is a specific kind of dependence relation between variables.

When we give causal explanations, we appeal to a kind of dependence that, in principle, we can use to manipulate the explanandum variable. The qualifier "in principle" needs careful specification. We may not in fact be in a position to change the value of some causal variable. We regard the mass of

the Earth as relevant to the causal explanation of such features of our Moon's orbit as its period, but we cannot perform experiments in which we change the mass of the Earth and observe changes in period of the Moon's orbit. What makes such an explanation causal nonetheless, according to Woodward, is that we provide information about the relationship between Earth mass and Moon period that allows us to answer *what-if-things-had-been-different* questions (Woodward, 2003, 11): If we were somehow to increase the mass of the Earth by 10 percent, how would the period of the moon's orbit change? What-if-things-had-been-different questions ask about *counterfactual* dependence relations.

Counterfactual conditionals are statements of the form 'If p were the case, then q' (D. Lewis, 1973). We often regard such claims as having clear meanings and readily determinable truth values. ('If you had called five minutes earlier, I would not have been able to answer the phone.') In other cases it is not clear whether such a claim is true or false because it is not clear to just what kind of situation the antecedent refers. ('If the ski resorts of Colorado had the climate of the Florida beaches, you could ski in your bathing suit.' Is this false because having the climate of Florida would rule out the presence of snow? But the antecedent seems to require that under the proposed scenario there are still *ski resorts*, so perhaps we should suppose that the freezing temperature of water would be much higher? But that would seem to necessitate a completely altered human physiology – lest our blood freeze – among other things …) The counterfactuals contemplated in what-if-things-had-been-different questions related to causal claims need to be of the former kind if causal claims are to have clear meanings.

When we cite a relation of causal dependence to explain why some variable takes a certain value, the related counterfactuals describe how the value of that variable would be different were we to *intervene* on the causal variable so as to set it to any of a range of possible values. Suppose that we specify 'smoking cigarettes causes heart disease' to mean that increasing the *number of cigarettes a person smokes per week* increases the *probability that the person will at some point develop heart disease*. This claim entails that, were we to somehow increase the cigarette consumption of a group of people, this would be an effective means to increase the rate of heart disease among them. This would be a terrible thing to do, and no one ought to carry out this manipulation. But this relationship between the manipulation and its terrible consequences is what the causal claim in question expresses.

Woodward's account thus draws a strong connection between causal explanation, experimentation, and control. As he states, "one ought to be able to associate with any successful explanation a hypothetical or counterfactual experiment that shows us that and how manipulation of the factors mentioned in the explanation ...would be a way of manipulating or altering the phenomenon explained" (Woodward, 2003, 11). Such manipulation takes place by way of *intervening* on the causal variable. The term 'intervention' has a technical meaning for Woodward. Through a detailed discussion he provides conditions intended to ensure that something counts as an intervention on variable X with respect to Y only if it changes Y *only* as a result of the causal influence of X on Y (Woodward, 2003, 94–104). This is necessary to avoid problems arising from cases where, for example, curare is administered by means of a curare tipped bullet shot into a subject's heart (thus producing death, but not from curare).

Explanation, then, is "a matter of exhibiting systematic patterns of counterfactual dependence" (Woodward, 2003, 191). A causal explanation of the value of variable Y that cites the value of variable X in the explanans exploits a systematic relationship, expressible by a generalization G, between two variables X and Y. G must describe how changes to the value of X would make a difference to the value of Y (it must be a *change-relating* generalization). Woodward notes that this systematic relationship might take the form of a law, but need not. Rather than the CL model's requirement that explanations must cite laws, Woodward's account requires reference to a relationship that is *invariant* across some range of conditions. Specifically, it must be invariant (i.e., hold true) across a range of possible interventions on X that, according to G, yield distinct values of Y.

Some change-relating generalizations are more invariant than others, though all must exceed some minimal threshold of invariance to count as explanatory. Suppose we determine the lethal dose of curare for a certain size monkey. We can express the relationship between curare dosage and mortality with a change-relating generalization G_1 which could then be invoked to explain the deaths of monkeys wounded by arrows delivering sufficiently strong doses of curare or the survival of some monkeys struck with arrows not thus prepared. But the same intervention on curare dosage might not be sufficient to bring about the death of an elephant. Presumably, no dose of curare would suffice to kill a tree. Our generalization G_1 does exhibit some invariance. It would (we might suppose) apply equally well

if the dose were administered with a syringe instead of an arrow, for example. A generalization that describes how curare interferes with the communication between motor neurons and muscles would have a greater degree of invariance. We would see this mechanism at work even in cases (like the elephant) where the dosage was insufficient to produce death. Woodward argues that the *depth* of a causal explanation corresponds to the degree of invariance of the generalization that it cites. (See, however, Strevens, 2007 for a criticism of this claim and Strevens, 2008 for an alternative account.)

Circularity?

According to Woodward, to say that variable X is a cause of variable Y is to say that intervening on X makes a difference to the value of Y. But to know whether the process that sets the value of X constitutes an intervention relative to Y requires sufficient causal knowledge to establish that the intervention changed the value of Y only through its effect on X and not through other causal connections to Y. This raises the worry that Woodward's account of causation is circular: In explaining causation, Woodward uses the very notion he is trying to explain!

Woodward's reply emphasizes two points. First, the account of causation he is offering is intended not as a metaphysical theory that shows how causal relations can be reduced to more basic ontological categories, but rather as a theory that clarifies the *meaning* of causal claims. Second, although determining whether a given means for fixing the value of X is an intervention with respect to Y does require some causal knowledge, it does not require knowledge of the causal relationship between X and Y. Thus, the circularity involved is not a vicious circularity. (Whether this response suffices is questioned in de Regt, 2004.)

11.5.3 Mechanistic explanation

The advancement of our understanding from simply knowing that hunters kill monkeys with curare-poisoned arrows to understanding the mechanism by which curare induces death exemplifies a common pattern of progress in many areas of science, particularly the life sciences. Let us finally consider an account of this kind of advance in our understanding that emphasizes the role of *mechanisms* in such explanations.

We read above a quotation from Carl Craver and Lindley Darden describing the mechanism whereby curare poisoning produces death. Craver and Darden, along with Peter Machamer (henceforth, MDC), argue that a significant aspect of scientific practice, at least in neurobiology and molecular biology, consists of the search for mechanisms and the explanation of phenomena by appealing to mechanisms (Machamer, Darden, & Craver, 2000). (MDC leave as an open question how widely their account of mechanisms applies, although they suggest a broad range of disciplines exemplify the pattern.) Moreover, they hold that "To give a description of a mechanism for a phenomenon is to explain that phenomenon, i.e., to explain how it was produced" (2000, 3).

MDC offer the following account of what such a description must include: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (2000, 3).

Note that on the MDC account an arrangement of entities does not suffice to compose a mechanism; one also needs *activities*. The mechanism whereby curare poisoning produces death includes things like tubocurarine chloride molecules and ACh receptors, but to achieve its final state the mechanism must include the *binding* of the former to the latter, which is how the tubocurarine chloride molecules *block* the ACh neurotransmitter and *paralyze* the muscle, ultimately *asphyxiating* the organism. MDC write "Activities are the producers of change. Entities are the things that engage in activities" (2000, 3). How the phenomenon is produced depends on how entities and activities are organized. A competing account of mechanisms by Stuart Glennan excludes activities from its ontology. On Glennan's account the mechanism for some behavior produces that behavior by "the interaction of a number of parts according to direct causal laws" (Glennan, 1996, 52).

MDC motivate inclusion of activities in their ontology by appealing to the kinds of things scientists deploy in their explanations: "There are kinds of changing just as there are kinds of entities. These different kinds are recognized by science and are basic to the ways that things work" (2000, 5). Moreover, describing how activities produce the termination conditions of the mechanism is required for mechanistic explanations to elucidate or make intelligible the phenomenon described by the explanandum (2000, 20–21). According to Craver, the reason that the HH equations for conductance changes in the action potential did not provide explanations of those changes

is that HH did not have evidence that the terms in those equations corresponded to the elements – both entities and activities – of the mechanism that produces those phenomena.

11.6 Conclusion

Scientists investigate systems, seeking to explain why they behave as they do. Many regard the understanding that such explanations yield as the principal benefit of the pursuit of scientific investigation. Thus it is no surprise that philosophers of science have devoted so much attention to attempting to understand scientific explanation. That the CL account is not adequate seems to constitute the one point of consensus in this discussion. Above the grave of the CL model a profusion of alternatives have blossomed, however, and opinions remain divided over the merits of these contenders. That we should not be able to settle on a single account of what constitutes an acceptable scientific explanation is, of course, just what Achinstein's argument would seem to predict. If nothing else, however, the various accounts considered in this chapter have served to highlight aspects of scientific explanation – the unifying power of explanatory theories, the potential for manipulation that causal explanations illuminate, the elucidation gained from understanding the mechanism behind a phenomenon – that enable us to better appreciate the ways in which scientific knowledge enhances and deepens our experience of the world.

12 Values in science

12.1 Introduction

We have already discussed how the logical empiricists sought to develop a *scientific philosophy* that would enable investigators to reason in a manner that would be independent of their commitments to ideologies or political causes.[1] Philosophers sometimes refer to this aspect of the logical empiricists' project as the ideal of a *value-free* science. But what are values?

Humans spend a lot of effort *evaluating* things, whether they be sneakers, songs, or senators. We can evaluate how honest a person is, how tasty a doughnut is, or how probable it is that an asteroid will collide with the Earth. One way in which such evaluations differ concerns what we value and how we value it. We value both honesty and tastiness, but in different ways. Valuing the continued survival of ourselves and other species, we might prefer that an asteroid not hit the Earth, but it would be odd to say that we value the probability of this happening: that would be something that we simply want to know. We might think of the category of things we value as those things that we take to be *good*, and the features of good things that give us reasons to value them (i.e., to regard them as good) we might call values. We should perhaps reserve the label *values* for those more general and

[1] This is not to say that the philosophers who pursued this project were not themselves motivated by political commitments of their own, for they were. Moreover, as the logical empiricist project developed it became apparent that 'neutral frameworks' in which the degree of confirmation afforded particular hypotheses by particular bodies of evidence could not be determined by criteria that were in any absolute sense a priori. The most celebrated of the logical empiricists working on theories of confirmation, Rudolf Carnap, acknowledged that only after a formal *language* had been chosen could one define an unambiguous measure of confirmation (Carnap, 1962), and that the choice of a language had to be made on the basis of either *pragmatic* criteria (Carnap, 1963, 980–983), or on the basis of what he later came to call *inductive intuitions* (Carnap, 1968).

fundamental kinds of features that give us such reasons. Suppose a pop song has a particular hook that you think helps make it a good pop song. *Having that particular hook* might not merit inclusion on a list of musical values for pop songs, but *being catchy* – the general attribute that the song achieves by having that particular hook – might well constitute such a value.

We can take things – including scientific theories, hypotheses, and methods – to be good in different *ways*. One distinction in particular is important here. We have already encountered the idea that scientists evaluate theories in terms of such characteristics as simplicity, accuracy, fruitfulness, and so on. Kuhn emphasized this point in his account of how theory choice could remain subject to incommensurability-driven underdetermination in spite of agreement amongst scientists about the importance of these values. It has become common since Kuhn's essay was published to distinguish *epistemic* from *non-epistemic* values (McMullin, 1982). The motivation for the distinction lies in a widely shared hunch: certain kinds of values have a legitimate role to play in scientific thinking, while others do not. A scientist may, for example, quite legitimately prefer one theory to another because it is more accurate. The scientist may not, however, legitimately prefer one theory to another because it reinforces his political opinions. We will see that this hunch must be qualified and clarified if it is to have any chance of being upheld.

Values, then, play a role in guiding decisions.What roles should values play in decisions that are made in the course of scientific inquiry?

We may start with some useful distinctions between different kinds of value-freedom that science might exhibit. Hugh Lacey distinguishes the following claims (Lacey, 1999):[2]

1. *Impartiality*: The only values that figure among the grounds for accepting and rejecting theories in science are epistemic[3] ones.
2. *Neutrality*: Accepting a scientific theory implies no commitments regarding non-epistemic values; neither does such acceptance undermine or support the holding of particular non-epistemic values.
3. *Autonomy*: Scientific inquiry is best carried out in a way that is free from 'outside interference' by social or political values or forces.

[2] These are approximations to Lacey's formulations. See (Lacey, 1999, ch. 10) for his more detailed statements.

[3] Lacey uses the term 'cognitive' here. Douglas distinguishes cognitive values from epistemic criteria of adequacy (Douglas, 2009, 93–94).

12.2 Decisions about research directions and methods

Not much reflection is needed to see that non-epistemic values must play *some* role in science. For example, the question of what kinds of scientific investigations are worth conducting depends significantly on value judgments about ethics and about what is socially beneficial. Given the finite resources available for the pursuit of scientific investigation, decisions about what kinds of scientific projects deserve support depend on judgments about the value of the knowledge those projects might yield. To some extent, funding agencies that provide support for scientific research depend on the expert judgments of scientific peers to judge, not just the likelihood that a specific proposed project will succeed, but also that the aims of the project are worth pursuing. At a more coarse-grained level of judgment, decisions about whether and how much to support broad categories of knowledge-pursuit, whether it is the treatment of juvenile diabetes or the study of the conditions in the early universe, are made by funding agencies themselves. For governmental agencies, some choices may even be directed or constrained by an elected legislature, for example by the differential allocation of funds to agencies pursuing varied research aims.

The use of public funds to pay for the costs of scientific research began in the nineteenth century.[4] Before then, investigators had either to spend their own money or convince other individuals to spend theirs. The increasing involvement of government in supporting science has coincided with the development of increasingly costly scientific undertakings, some of which, such as the Large Hadron Collider at the CERN laboratory in Geneva, Switzerland, are too expensive for any single nation to fund. Of course, funding for scientific research continues to come from private individuals or organizations, but public funding remains dominant, in spite of recent declines. According to the 2012 *Science and Engineering Indicators* report by the US National Science Board, the percentage of the total funding (US$54.9 billion) for Science and Engineering Research and Development provided by the US Federal Government in 2009 was 59 percent (a somewhat lower figure than in the 1970s and 1980s) (National Science Board, 2012).

[4] Laura Snyder tells the history of this development in England, and the philosophical issues that surrounded it, in her engaging and illuminating book, *The philosophical breakfast club* (Snyder, 2011).

The involvement of legislative bodies in determining levels of support for scientific research entangles the scientific community with broader political negotiations. Although scientists have never operated in complete isolation from their political and social contexts, the era of government dominance in the funding of research brings with it increased public attention to and involvement in the directions pursued by researchers. The results may yield occasions for either celebration or condemnation depending on your own political convictions. The US Congress in 1996 removed from the budget of the Center for Disease Control and Prevention (CDC) exactly the amount of money (US$2.6 million) that the CDC's National Center for Injury Prevention and Control spent on research into injuries caused by firearms, while adding to the CDC's appropriation a prohibition on use of its funding "to advocate or promote gun control." This restriction was later extended to all agencies within the US Department of Health and Human Services, and the effect has been to greatly diminish the amount of research carried out on the patterns and causes of gun-related injuries and deaths and on the effectiveness of measures to reduce them. The National Rifle Association, an organization that lobbies against restrictions on firearms, criticized federally funded scientists who presented the results of their research at meetings where others advocated gun control measures, and eventually the research stopped (Kellermann & Rivara, 2013; Davidson, 2013).[5]

Finally, citizens have influenced the ways in which research is carried out based on their ethical values by acting to prohibit the use of research methods that they consider unethical. One point of ongoing debate concerns the use of non-human primates as model organisms in biomedical research. Researchers using them note that their genetic and physiological similarity to human beings makes non-human primates ideal for studies in which it would be impractical or unethical to use human subjects. Critics of this practice point to the same resemblances as a reason why we should regard such uses as morally problematic or even repugnant. A number of European countries, as well as New Zealand, have banned the use of great apes in biomedical research, and in the US, the Director of the National Institutes of Health announced in December 2011 that the NIH would suspend new grants

[5] President Barack Obama issued a plan to reduce gun violence in January 2013 that included a call for the resumption of federally funded research into the "causes and prevention of gun violence" (Davidson, 2013).

for biomedical and behavioral research on chimpanzees (Gorman, 2011). Subsequently the NIH has undertaken to "retire" most of the chimpanzees it holds for research purposes to a sanctuary where they would be ineligible for use as research subjects and to both reduce the number of studies using chimps and raise standards for chimpanzee living conditions in laboratories where they are used (Gorman, 2013). A report commissioned by the Institute of Medicine that found that "most current use of chimpanzees for biomedical research is unnecessary" provided the basis for these changes (Committee on the Use of Chimpanzees in Biomedical and Behavioral Research, 2011, 4).

Clearly, in cases such as these, scientists and other interested people act to influence the direction of research as well as the methods used to pursue that research, based on their ethical and social values. Scientists and trained ethicists may influence government agencies through participation in a committee that either a government body or an independent organization has called upon to prepare an advisory report. Citizens who are simply interested in an issue, whether pursuing scientific research as a profession or not, may also exert pressure on policy-makers to restrict or encourage research on a particular topic and to change or restrict the methods that researchers may use. It seems entirely reasonable in a democracy that scientific research be subject to such influences, casting considerable doubt upon the *autonomy* version of the value-free ideal.

12.3 The argument from inductive risk

Regardless of one's stand regarding questions about the direction of science or the moral acceptability of certain research techniques, that non-epistemic values should play a role in these debates seems appropriate and does not raise any serious concerns about the objectivity or rationality of scientific inquiry or its results. Whatever considerations might have affected the selection of the research question, and whatever the source and nature of the constraints imposed on permissible methods in pursuit of an answer to that question, once the scientist is engaged in the collection and analysis of data, non-epistemic values have no further role to play. Determining the answer to the research question that is best supported by the data generated using the selected method requires no further consideration of what is permissible or desirable, morally or otherwise.

Or so it might seem.

Even when they apply them correctly, the methods scientists use to make inferences from data may lead them to draw conclusions *in error*. One argument for the relevance of non-epistemic values to the core tasks of scientific reasoning points to this feature and to the fact that not all errors that a scientist might make are equally serious. Insofar as non-epistemic values figure into our differential attitudes toward such errors, they affect our choice of methods for drawing conclusions from data, and hence the conclusions themselves.

Somewhat more systematically, the *argument from inductive risk*, as it is sometimes called, runs as follows:

1. In order to draw a conclusion from data, the scientist must make a choice regarding the method to be used in reaching a conclusion.
2. The probability of making different kinds of errors depends on the choice of method.
3. Non-epistemic values may legitimately influence the scientist's preference regarding different kinds of errors.
4. Therefore, non-epistemic values may legitimately influence the scientist's choice of methods for drawing conclusions from data.
5. Thus, non-epistemic values may legitimately make a difference to the conclusion the scientist draws from data.

12.3.1 Example: formaldehyde again

The argument from inductive risk may be best appreciated if we consider a concrete example of statistical hypothesis testing in which non-epistemic values are clearly at stake in some way. In Chapter 9 we discussed how scientists used statistical tests to investigate the carcinogenicity of formaldehyde.

We can start with a simplified reconstruction of the inferences and decisions involved in the formaldehyde case so that we may consider how the argument from inductive risk might apply in this case. (We will subsequently note the problems with relying on this simplification.)

We might think of the tests performed on rats and mice in order to evaluate carcinogenicity as tests of the *null* hypothesis that the inhalation of formaldehyde does not increase the risk of cancer. (Recall our discussion of such tests in Chapter 9, Section 9.3.) The *alternative* hypothesis is then that formaldehyde does increase the risk of cancer by some amount. In order to

specify a Neyman–Pearson test, we need to specify the *critical region*, i.e., those values of the test statistic for which the test rejects the null and accepts the alternative hypothesis. Assuming that the test statistic has been determined, the choice of critical region determines the Type I error probability for the test, i.e., the probability that the test will reject the null hypothesis, assuming that it is true. So in order to choose the critical region, we need to decide what Type I error rate we wish to achieve, i.e., what is the maximum probability of rejecting the null hypothesis when the null is true that we will tolerate? In answering this question, of course, we must keep in mind that the lower we set the Type I error rate (the significance level) of the test, the higher the Type II error rate for any given member of the set of possibilities that constitutes the alternative hypothesis. Recall that a Type II error consists of not rejecting the null when the alternative hypothesis is true.

In the formaldehyde case, then, a Type I error would be to declare formaldehyde to increase the risk of cancer when it does not, while a Type II error would be to fail to declare formaldehyde to increase the risk of cancer when it does. The argument from inductive risk would draw our attention to the fact that our decision regarding the appropriate method of testing for these hypotheses will involve a choice about how to balance these error probabilities. That decision in turn may quite appropriately draw upon our differential attitudes regarding the consequences of one error as opposed to another. Since what is at stake is the health of human beings (especially those employed in certain industries), we are well justified if we design the test to have a significance level that is relatively high. However, we must also consider the negative economic impacts on certain businesses and their employees that might result from a Type II error.

We certainly would not want to demand a Type I error probability as low as that required by particle physicists for declaring the discovery of a new particle, which you may recall from our discussion in Chapter 9 is (for at least some cases) about 3×10^{-7}. Particle physicists can afford to be very conservative about Type I errors, because the consequences of failing to report the discovery of a new particle when one is present would not be so serious – the experiment would simply have to run a bit longer to have sufficient data to meet the statistical threshold for making an announcement. (By contrast, erroneously announcing a discovery could be disastrously embarrassing for particle physicists and undermine public support for their quite expensive, publicly funded research.) But, if formaldehyde exposure really does increase the risk

of cancer and scientific investigations fail to detect it, then dire consequences may result. Government regulatory agencies responsible for workplace safety standards would not have a basis for eliminating or limiting exposure to a carcinogenic substance, leaving people working in industries in which prolonged exposure to formaldehyde is prevalent vulnerable to high rates of potentially lethal cancers that could be prevented by appropriate regulation.

We should pause to note an important complication, however. Testing for carcinogenicity of formaldehyde follows the typical procedure of setting a point null hypothesis (the increased risk of cancer equals zero) against a compound alternative (the increased risk is greater than zero). This means that although a Type I error can be described simply, Type II errors vary widely in their consequences. Failing to reject the null even though formaldehyde exposure increases the risk of cancer by 1000% commits a Type II error. So does failing to reject the null when the increase is 0.01%. The severity of the consequences of these two errors for human health differs greatly.

From the perspective of the argument from inductive risk, this complication – that different Type II errors will differ in their potential for causing harm – simply reinforces a crucial premise: That the investigator may, on non-epistemic grounds, judge some possible errors as more undesirable than others. Choosing a significance level for a test that would have a good chance of detecting even as small an increase in risk as 0.01% may well, given limitations on the affordable sample size, turn out to be achievable only by having a test that will quite frequently reject the null, even when it is true. A rejection of the null based on such a test would not constitute strong evidence that the null is false, according to the severity principle of Chapter 9, Section 9.4.1. Allowing people to be exposed to substances that increase the risk of cancer always (all else being equal) increases the probability of a terrible outcome for someone. Choosing instead to test the null in a way that has a lower Type I error rate at the expense of being insensitive to such very small increases in risk *might* indicate that the investigator has made a decision that the importance of avoiding such a relatively low-probability outcome is trumped by the epistemic value of conducting a test that has the potential to produce strong evidence either for or against hypotheses under consideration.

In his version of the argument from inductive risk, Richard Rudner identifies the crucial premise of the argument as the claim that scientists *accept* hypotheses on the basis of the tests that they perform. Since those tests do not in general yield conclusive proof, the decision to accept a hypothesis depends

on a judgment that the evidence supporting the hypothesis is *sufficient* to make such acceptance a reasonable choice. As long as there are practical *consequences* of such acceptance, value judgments regarding those consequences will be relevant for determining how good the evidence must be in order to be sufficient in this sense (Rudner, 1953).

12.3.2 Skeptical responses to the argument from inductive risk

To this, however, one might object that in doing so the scientist assumes a task that is not properly hers. As a scientist, she ought to report the strength of evidence regarding relevant hypotheses, and not presume to decide whether such evidence is sufficient to accept any particular hypothesis for purposes of practical decision-making.

Richard Jeffrey argues this point in an article published directly in response to Rudner's essay. The scientist ought not to accept hypotheses, according to Jeffrey, because she could not make a decision about acceptance in a way that would be optimal for each problem for which that hypothesis would be relevant (Jeffrey, 1956). Jeffrey offers an alternative view of the task of the scientist, according to which she ought only to report a probability for whichever hypothesis is under investigation, leaving it to others to use that probability assessment, in combination with their own values, as a basis for practical decisions. (Such a probability would have to be some sort of Bayesian probability, of course, since frequentist probabilities do not apply to general hypotheses.)

Jeffrey notes that Rudner had already anticipated such a response and argued that such a move would not excuse the scientist from the task of accepting hypotheses because the claim that the probability of a given hypothesis is p is itself a hypothesis. But, replies Jeffrey, the scientist should not *accept* this kind of hypothesis either! (Jeffrey, 1956, 246). Unfortunately, he offers no explanation of how to distinguish reporting a probability statement to be made use of in further practical decisions from accepting that same probability statement.

Moreover, even if one supposes that in reporting the outcomes of an investigation the scientist can in this way avoid accepting the hypotheses on which they report, scientists must, as Duhem emphasized, *rely* on hypotheses in the design of experiments, operation of instruments, and analysis of data, which surely counts as a kind of acceptance. Thus, even were it

possible to decline accepting hypotheses that are under investigation in a given experiment, one must accept other hypotheses in order to carry out an experiment. Considerations that enter into such decisions may rightly include not only the sufficiency of the experimental arrangement for the intended investigation but also the safety, affordability, or ethical acceptability of the proposed procedure. We have been supposing that we can neatly segregate the entry of non-epistemic values into considerations of the moral acceptability of experimental techniques from questions regarding the analysis and interpretation of data, but the dependence of the latter on previously tested or at least testable scientific hypotheses casts doubt on how well we can maintain that separation.

Another response to Rudner's argument acknowledges that scientists do accept hypotheses, but denies Rudner's conclusion that scientists must make non-epistemic value judgments in doing so. According to Isaac Levi, Rudner's argument assumes a behavioristic understanding of statistical inference that attempts to reduce scientific reasoning to "technology and policy making" (Levi, 1962, 47). Levi offers an alternative account according to which the goal of hypothesis testing is sometimes to replace doubt with belief, and in some such cases the scientist seeks "the truth and nothing but the truth" (Levi, 1962, 49). In such cases, the investigator must not differentiate between the seriousness of different errors, but treat all errors as equally serious.

Levi directs his arguments primarily not at the argument from inductive risk, but at the *behavioristic* interpretation of hypothesis testing in science, according to which the act of accepting a hypothesis must be given an interpretation as a decision to take some practical action. Yet he clearly considers the two issues to go hand-in-hand. Values are necessarily involved in the inferences that scientists draw from data if and only if all such inferences must be interpreted in behavioristic terms (Levi, 1962, 50).

Such a linkage between the role of non-epistemic values in scientific reasoning and the behavioristic interpretation of hypothesis testing becomes understandable in the context of the literature of the mid twentieth century to which Levi was responding. C. West Churchman had linked scientific hypothesis testing with practical action in his version of the argument from inductive risk (Churchman, 1948). Churchman insisted that "every scientific hypothesis is considered to be a possible course of action for accomplishing a certain end, or set of ends. ... Statements like 'we merely want to find out so-and-so' represent pragmatically incomplete formulations" (Churchman, 1948, 259).

Critics of the argument from inductive risk apparently concede that within the behavioristic interpretation of hypothesis testing, the argument from inductive risk validly supports the conclusion that non-epistemic values may legitimately influence scientific reasoning. In Chapter 9 we discussed an alternative *evidential* approach that regards statistical hypothesis testing as a framework for the evaluation of the evidence supporting inferences from data to hypotheses under test. Is the argument from inductive risk still valid if we employ this evidential interpretation?

To exclude non-epistemic values from playing a role in scientific reasoning is, in practice, to suppose that we could impose a division of labor between scientists, responsible for evaluating the evidence regarding some scientific hypotheses, and policy-makers, responsible for weighing the potential costs and benefits of various policies in light of the evidence relevant to the suitability of those policies for attaining the particular public goods at issue. As long as we stick to our simplistic conception of how testing works in the testing of formaldehyde and similar cases, such a division of labor might seem possible, but actual hypothesis testing involves numerous kinds of judgments that complicate this simple picture.

In a discussion of the very same formaldehyde example, Deborah Mayo articulates a number of questions that must be answered in the course of utilizing scientific data from both epidemiological studies and animal model experimental studies to inform policy decisions. These questions concern "inference options," and have "no unequivocal scientific answers," but because "these choices have policy implications, they are intertwined with policy" (Mayo, 1991, 257). Such questions include what weight to give to studies yielding different results, what significance level to consider as indicating a positive result, how to extrapolate from doses used in experimental studies to doses that are encountered in policy-relevant situations, and how to extrapolate from the dose response of model organisms to the response of humans (Mayo, 1991, 258). Making appropriate decisions on these matters requires significant scientific expertise and judgment, and thus cannot be left to policy-makers. At the same time, different choices have clear policy implications insofar as they raise or lower the likelihood that, for example, a substance such as formaldehyde will be found to pose a significant risk.

Thus, at least in contexts in which scientists assess evidence that is relevant for policy-making, non-epistemic values ought to inform decisions concerning inference options. That entanglement of non-epistemic values in scientific reasoning does not depend on a behavioristic interpretation of

hypothesis testing. We may take the scientists studying the carcinogenicity of formaldehyde to be using statistical methods to evaluate the evidence supporting a variety of possible inferences from the data, as the evidential interpretation requires. The argument from inductive risk requires only that scientists' choices regarding inference options make a difference to their conclusions regarding evidence, and that those choices are based in part on the potential influence that such conclusions may have on policy-makers' decisions. Even if there are contexts in which scientists do, as Levi puts it, seek "the truth and nothing but the truth" (a premise that pragmatically oriented philosophers such as Churchman would deny) it does not follow from this that we should as a general rule regard the entry of non-epistemic values into scientific reasoning as *illegitimate* or *unscientific*.

12.3.3 Direct and indirect roles for values

You would be right to be dissatisfied with leaving the matter there, however. Supposing we accept that non-epistemic values *may* play a legitimate role in scientific reasoning, we still face the question of when they *ought* and *ought not* do so. That scientists should seek to understand how the world really is in a way that is not influenced by how they wish the world to be retains its intuitive appeal.

Moreover, allowing values to enter into scientific reasoning willy-nilly might exacerbate current threats to the integrity of scientific research. To cite just one example, the role of pharmaceutical companies in funding research into the safety and effectiveness of their own products might lead to bias in the results of such research (Lemmens & Waring, 2006). Researchers have found a statistically significant association between industry funding and pro-industry results (Bekelman, Li, & Gross, 2003). Although investigators may certainly maintain the independence of their results from the source of their funding, the potential for bias is significant (Wilholt, 2009). Investigators clearly should not count the potential for profit as a reason influencing their judgments regarding the effectiveness and safety of a treatment. Given that we allow some role for non-epistemic values in science, on what basis do we exclude cases such as this?

To answer this question, Heather Douglas argues, we need to distinguish distinct *roles* that values might play in science, which she calls *direct* and *indirect*. In their direct role, values "act as reasons in themselves to accept

a claim, providing direct motivation for the adoption of a theory" (Douglas, 2009, 96). When acting in their indirect role, values "weigh the importance of uncertainty about the claim, helping us to decide what should count as *sufficient* evidence for the claim" (Douglas, 2009, 96; emphasis in original). She goes on to observe that in their direct role, "values determine our decisions in and of themselves, acting as stand-alone reasons to motivate our choices. ... In this direct role, uncertainty is irrelevant to the importance of the value in the judgment" (Douglas, 2009).

(In characterizing Douglas's position, note that I have referred to these as roles for values in general and not for non-epistemic values alone. Douglas rejects the distinction between epistemic and non-epistemic values. According to her, what are typically considered epistemic values are really *criteria* of adequacy that scientific claims must satisfy, rather than values. For her, not only is the distinction between different roles for values in science more important than the distinction between epistemic and non-epistemic values, it completely replaces that distinction. Daniel Steel argues that Douglas's distinction between roles does not succeed in supplanting this distinction between types of values. He defends the view that epistemic values are those that promote the attainment of truth (Steel, 2010). We will not pursue this aspect of the debate further; we may read Douglas's references to 'values' as indicating 'non-epistemic values' in order to focus on the question of what roles such values may legitimately play in science.)

Values may play indirect roles throughout science without necessarily threatening scientific integrity, according to Douglas, but their direct roles must be limited if we are to maintain "the integrity of the scientific process" (Douglas, 2009, 96). We have already discussed some legitimate direct roles that values may play, such as in decisions concerning what questions to investigate or what research methods to use. Even here, Douglas notes, we need to scrutinize how values influence decisions, since they could, for example, lead one to choose a method of inquiry that "predetermines (or substantially restricts) the outcome of a study" (Douglas, 2009, 100). One could select for one's study of the effects of formaldehyde on rats only rats who are already in extremely frail health, for example. Or, underlying values-based beliefs might bias an individual or an entire community toward considering only certain kinds of questions. Some critics, for example, have alleged that modern medical science's emphasis on intervention and control – what Hugh Lacey (2005) has in another context called the "modern

valuation of control" – biases it towards research into problems such as cancer that emphasize pharmacological and surgical treatments at the expense of investigations into environmental factors that might be causing increased rates of cancer (Epstein, 1998; Kourany, 2010, 122–125).

Douglas insists, however, that any direct role for values in science must be restricted to these "early stages" of research (choosing a question to investigate, choosing a method of investigation). "Once the study is underway," values should play no direct role, except in "unusual circumstances when the scientist suddenly realizes that additional direct value considerations need to be addressed" (Douglas, 2009, 101). In a clinical study of some potential medication, for example, one might find that patients in the treatment group are suffering unexpected adverse effects, requiring one to terminate or modify the study for the good of the subjects. Alternatively, one might find that patients in the treatment group are so obviously benefiting in a way that patients in the control group are not that moral demands compel the assignment of all patients to the treatment group (Lilford & Jackson, 1995; Worrall, 2008). Aside from such situations (which are increasingly accounted for in advance in the development of sophisticated research plans), once the investigation is under way, "values must be restricted to an indirect role" (Douglas, 2009, 102).

The argument against allowing a direct role for values in the "later stages of science" is that doing so would allow inferences from scientific data to be influenced or even determined by the values of the investigator. Scientists could "reject data if they did not like it, if, for example, the data went against a favorite theory," or they "could select an interpretation of the evidence because they preferred it cognitively or socially, even if the evidence did not support such an interpretation," or they could reject a theory they dislike "regardless of the evidence supporting it." By doing so, "we would be causing harm to the very value we place in science," i.e., its ability to provide us with a "more reliable, more true understanding of the world" (Douglas, 2009, 102).

It would not harm this value, however, for scientists to allow values an *indirect* role in the later stages of inquiry, such as determining the level of significance that must be achieved to reject a null hypothesis. Douglas thus upholds the argument from inductive risk (though with a narrowed scope for its conclusion): scientists may legitimately consider the different values attached to different possible errors in deciding what counts as sufficient evidence for accepting a hypothesis. Allowing a direct role for values in the

drawing of conclusions regarding the questions under investigation, however, would allow human preferences to supplant the role of empirical data in scientific inquiry.

Douglas's distinction between direct and indirect roles marks a noteworthy contribution to our examination of this issue, yet it does not seem to address fully the question of when non-epistemic values may or may not enter into scientific reasoning. As Kevin Elliott has pointed out, the terms in which Douglas draws the distinction are somewhat unclear. What does it mean for something to function, for example, as a "stand-alone reason" (Elliott, 2011, 306–310)? Moreover, it seems that the problems that arise from values entering into a direct role in drawing conclusions from an investigation can result from indirect roles as well. The investigator who wishes either to cling to a theory she does not want to give up or reject a theory she loathes may simply let those preferences serve as a basis for criteria of sufficiency that, by being permissive or restrictive, make her preferred outcome more likely (Wilholt, 2009).

12.4 Values and the objectivity of science

Our discussion thus far has concerned the negative or positive roles that values might play in influencing the decisions of scientists engaged in research. A somewhat different set of questions concerns how values shape the conduct of scientific work. Does the achievement of the aims of science require commitment to certain values, and if so, which ones? Should scientific research be organized and institutionalized according to certain values? How might broader social values contribute to or erode the objectivity of science?

12.4.1 Merton on the ethos of science

The sociologist of science Robert K. Merton,[6] writing in 1942, sought to describe what he called "the ethos of science": "that affectively toned complex

[6] When Merton published his doctoral dissertation 'Social and cultural contexts of science' in 1938 (Merton, 1938) the sociology of science did not really exist as an established discipline within Anglophone sociology. A lively field bristling with theoretical and empirical debates, contemporary sociology of science has developed in ways that the erudite Merton could not have anticipated in those early years when, as he recalled many years later, "An abundance of [sociological] monographs dealt with the juvenile delinquent, the hobo and saleslady, the professional thief and the professional beggar, but not one dealt with the professional scientist" (Merton, 1973, 173).

of values and norms which is held to be binding on the man of science" (Merton, 1973, 268–269). Merton posits the *institutional goal* of science to be "the extension of certified knowledge," an aim to be achieved through technical methods adapted to the attainment of "empirically confirmed and logically consistent statements of regularities" (Merton, 1973, 270). From the institutional goal and technical methods, Merton claims, derive four sets of *institutional imperatives* that function as "moral as well as technical prescriptions in science": universalism, communism, disinterestedness, and organized skepticism (Merton, 1973, 270). That these norms function as imperatives within science he concludes not from the observation of their universal satisfaction by scientists – on the contrary, scientists rather frequently and sometimes systematically violate all four of them – but from the ways in which violations are rationalized by the violators and condemned by the defenders of these norms.

Universalism, according to Merton, finds expression in the requirement that scientific claims should be evaluated according to "preestablished impersonal criteria" rather than by consideration of "personal or social attributes of their protagonist" (Merton, 1973, 270). Moreover, this norm requires that one's prospects for a career in science should be determined by one's abilities and accomplishments rather than by one's station in society. "To restrict scientific careers on grounds other than lack of competence is to prejudice the furtherance of knowledge" (Merton, 1973, 272). Merton had, at the time of his writing, ample opportunity to observe widespread and systematic violations of this norm. German national socialists had insisted, for example, that only Aryans could produce 'good' science, and as a consequence had expelled Jews from their academic institutions. (Merton did not at the time discuss the widespread exclusion of African-Americans from science in the United States and elsewhere, or the obstacles that women faced in the pursuit of scientific careers.) Merton argued that systematic violations of universalism were bound to arise because "science is part of a larger social structure with which it is not always integrated," and when the broader culture surrounding science develops anti-universalistic tendencies, this will put the ethos of science under considerable strain, leading to rationalizations seeking to show that the members of some group despised in the broader culture are incapable or ill-disposed towards producing worthwhile scientific contributions (Merton, 1973, 271).

The norm of *communism* holds that the results of scientific investigation "are a product of social collaboration and are assigned to the community. ... Property rights in science are whittled down to a bare minimum by the rationale of the scientific ethic" (Merton, 1973, 273). We might reward discovery by attaching a scientist's name to a result, but the estate of James Clerk Maxwell does not get to charge royalties whenever 'Maxwell's laws' of electromagnetism are invoked. The increasing role of private capital in the pursuit of potentially profitable scientific research has exerted considerable pressure on this norm in recent years, particularly in the United States, where laws protecting intellectual property are especially robust. (However, a recent effort by Myriad Genetics to patent a DNA sequence was rejected by the US Supreme Court in an important decision reached in June 2013.)

In his discussion of *disinterestedness* as a norm in science, Merton emphasizes that this is to be interpreted as a norm operating at the institutional level rather than as a description of the motivations of the individual scientist. However biased or partisan the individual scientist might be in his thinking about the subject of his investigations, the demand that his results be both testable and publicly available subjects the investigator to "the exacting scrutiny of fellow experts" (Merton, 1973, 276). This explains what Merton describes as the "virtual absence of fraud in the annals of science" (276). (It is not clear how well this assessment stands up in the present day.[7]) As a kind of corollary to the self-enforcement of the norm of disinterestedness *among* scientists, Merton notes in contrast that in relations between scientific experts and laypersons, the authority of science "can be and is appropriated for interested purposes, precisely because the laity is often in no position to distinguish spurious from genuine claims to such authority" (Merton, 1973, 277).

[7] There have been some scandals involving scientific fraud recently. The story of social psychologist Diederik Stapel at the University of Tilburg, who admitted to fabricating data in a significant number of publications is both dramatic and troubling (Bhattacharjee, 2013). Yet, in the spirit of Merton's emphasis on the self-policing of scientists to enforce disinterestedness, one can also visit the blog Retraction Watch, which tracks retracted publications "as a window into the scientific process." Although most of the retractions reported there are not instances of fraud, Retraction Watch does display the wide variety of ways in which scientists can – intentionally or unintentionally – wind up in a pickle.

Merton describes *organized skepticism* as "both a methodological and an institutional mandate" (Merton, 1973, 277) that periodically puts science into conflict with other institutions. The willingness of scientists to subject to critical examination beliefs that religious or political authorities regard as inviolable "appears to be the source of revolts against the so-called intrusion of science into other spheres." Such conflicts arise from the application of scientific inquiry to questions regarding which "institutionalized attitudes" have been established or when other institutions "extend their control over science" (Merton, 1973, 278).

Merton's analysis of these institutional norms makes clear the *sociological* function of these imperatives without delving into their *epistemological* aspects. More recent discussions among philosophers of science have sought to illuminate how these and other values might *contribute* to the objectivity of scientific knowledge.

12.4.2 Longino on contextual values

Helen Longino seeks to reconcile an active role for contextual values in science with scientific objectivity. No one doubts the importance of *constitutive values* in science, meaning values "generated from an understanding of the goals of science" (these might include what we have been calling "epistemic values," though we might also consider these to encompass Merton's norms). But calls for a "value free" science presumably mean to exclude "those group or individual preferences about what ought to be" that Longino calls *contextual values* (Longino, 1990, 4).

Longino's reconciliation begins with a claim about evidence: no fact or sentence describing a factual state of affairs serves as evidence for a particular hypothesis all on its own. Considered in isolation, facts are evidentially meaningless; they serve as evidence for particular hypotheses (meaning, for Longino, that scientists can *use* them to argue for such hypotheses) only in the context of particular sets of background beliefs. A crucial problem about evidence arises from this claim: to the extent that background beliefs used in evidential arguments are themselves susceptible to influence from contextual values, how can the objectivity of scientific inferences from evidence be secured?

Longino discusses a broader range of value-influences in science than that which emerges in discussion of inductive risk. The latter involves value

judgments concerning the costs and benefits of erroneous or correct inferences. Longino's discussion encompasses a range of ways in which contextual values "shape the knowledge" that emerges from research programs in science, sometimes involving subtle entanglements among moral and epistemological considerations. We will here focus on Longino's analysis of cases in which values influence judgments about what kinds of explanations are best for certain kinds of data. Such background assumptions often take the form of what Longino calls "explanatory models," which describe "the sorts of items that can figure in explanations of a given sort of phenomenon and of the relationships those items can be said to bear to the phenomena being explained" (Longino, 1990, 134).

In the sciences that deal with human behavior such value judgments have long vexed investigators, the more so because they often remain unstated. In her 1990 book, Longino describes what she then took to constitute the dominant explanatory model assumed in investigations into gender role behavior, the development of homosexuality, and sex differences in cognitive performance: The *linear hormonal model* posits certain types of behaviors as the outcome of a process in which exposure of the fetal organism to testosterone, estrogen, or their metabolites plays a causal role in the determination of the brain's structure and functioning, so that animals wind up with "male brains, programmed for characteristically male behavior, or female brains, programmed for characteristically female behavior" (Longino, 1990, 137). Longino argues that only in the context of this linear hormonal model do the kind of data gathered in studies of hormonal exposure and behavior provide evidence for causal hypotheses linking behavior to hormonal exposure.

Longino goes on to show how a different model might lead to a significantly different understanding of the behaviors being investigated. According to what Longino refers to as the *selectionist model*, the functions of the brain are carried out by groups of about 50 to 10,000 neurons, and connections between them can be strengthened or eliminated in the process of brain development. We need not concern ourselves with the details of the mechanism; the important point is that, as Longino speculates, such a model allows for a much greater potential role for the experience of the developing individual (including experience of the culturally influenced responses of others to physical appearance) in the explanation of patterns of behavior.

Neither the linear hormonal model nor the selectionist model, Longino explains, enjoys definitive direct empirical support. Both can be given plausible but inconclusive empirical and conceptual motivations. Neither does either model obviously suffer fatal empirical weaknesses. Moreover, such scientific research into the nexus of gender and behavior enters a broader culture of lay persons and policy-makers eager to learn 'what science has to say' about such issues as child-raising, gender differentials in academic fields and professions, gender relationships in the family, and sexual orientation, and often even more eager to exploit the latest results to defend their own value judgments and policy proposals. In such circumstances, where constitutive scientific values fail to single out an explanatory model that will enable the determination of causal explanations from otherwise ambiguous data, implicit appeals to contextual values can be expected to play an important role in filling this gap between data and evidence.

This returns us to our previous question, however. With contextual values playing such an important role in how we draw conclusions, it seems that the very possibility of achieving objectivity lies in peril. Advocates of a value-neutral ideal for science will see in this situation the conditions that give rise to 'bad science' and demand that we strip away any influence on our conclusions from contextual values. Longino agrees both that objectivity is an important aim of scientific inquiry, and that some ways in which contextual values influence scientific reasoning threaten that objectivity. But because she does not think that contextual values can be eliminated from influencing how the gap between data and evidence gets filled, she offers a different vision of how objectivity is to be secured: objectivity results from the *interaction* of investigators holding different, competing contextual values through social mechanisms that, in virtue of satisfying specifiable criteria, enable transformative criticism.

Longino's response deploys a certain notion of objectivity. One might think of the objectivity of science in terms of an "accurate description of the facts of the natural world as they are." To conceive of the objectivity of scientific knowledge in these terms is to embrace some version of realism about science, in the sense of making the truth of descriptions derived from scientific theory central to the aims of science. Longino's account of the objectivity of science employs a different notion, according to which "objectivity has to do

with modes of inquiry" (Longino, 1990, 62).[8] To describe science as objective in this methodological sense is to regard the development, acceptance, and rejection of scientific theories and hypotheses as guided by criteria that are "nonarbitrary and nonsubjective" (Longino, 1990, 62).

To see the extent to which scientific inquiry achieves methodological objectivity, according to Longino, we must consider not only the logical forms of argumentation that scientists instantiate, or the rules that they follow, but the *practices* of science, conceived of as implemented by social groups. Longino here draws on work by Marjorie Grene (1966, 1985) in the philosophy of science and Alasdair MacIntyre (1981) in ethics. Not only does the production of scientific knowledge require the cooperative and coordinated action of groups of researchers as a practical matter, but, as Grene emphasizes, skill in the practices that constitute the methods in any given discipline requires a kind of initiation into "the traditions, questions, mathematical and observational techniques, 'the sense of what to do next'," and this initiation must take place at the hands of someone who has already been through it and practiced the acquired skill within the community (Longino, 1990, 67). With Grene, Longino takes these socially situated practices to indicate that the sciences constitute something akin to what MacIntyre describes as *practices*. This term, MacIntyre explains, refers to "any coherent and complex form of socially established cooperative human activity through which goods internal to that form of activity are realized in the course of trying to achieve those standards of excellence which are appropriate to, and partially definitive of, that form of activity" (MacIntyre, 1981, 175).

Precisely what constitute the distinctive internal goods of science Longino declines to specify, considering such specification to be part of the evolving self-conception of science itself, an outcome of a socially complex and historically contextualized process (Longino, 1990, 19). Instead, Longino offers criteria by which to judge the social organization of science with regard

[8] As emphasized by K. Brad Wray, Longino rejects the traditional view of knowledge as requiring truth. She wishes to retain the normative function of the philosophically traditional concept of knowledge, and the distinction between knowledge and mere belief, but to extend the category of knowledge to include representations (including non-linguistic models of the sort discussed in Chapter 4) that are empirically adequate within some domain and sufficient to support both practical and investigative projects (Longino, 1994; Wray, 1999).

to its ability to achieve methodological objectivity. It is precisely the social character of science that enables the mitigation of subjective preferences, but not just any social organization will do (the pop music industry is also socially organized, after its own fashion). Science can only achieve objectivity if it is organized to promote "criticism from alternative points of view" and "the subjection of hypotheses and evidential reasoning to critical scrutiny. . . . A method of inquiry is objective to the degree that it permits *transformative* criticism" (Longino, 1990, 76; emphasis in original).

Longino specifies four criteria "necessary for achieving the transformative dimension of critical discourse" (Longino, 1990, 76). A scientific community must have:

1. "recognized avenues" for criticism that may be directed at all aspects of scientific reasoning (including methods and background assumptions). Scientific journals and professional meetings obviously fulfill this function.[9]
2. "shared standards" by which to evaluate both the arguments advanced by investigators and the criticisms directed at those arguments. Longino notes the similarity of her view of such standards and their importance with that articulated by Kuhn in his essay "Objectivity, value judgment, and theory choice," but emphasizes that "the open-ended and non consistent nature of these standards . . . allows for pluralism and for the continued presence, however subdued, of minority voices" (Longino, 1990, 77–78).
3. a capacity and habit of responding to critical discussion by changing its beliefs. Longino does not take this to require that investigators "recant" whenever criticism is directed at their data or assumptions. "Indeed, understanding is enhanced if they can defend their work against criticism" (Longino, 1990, 78). But it does require that "the assumptions that govern their group activities" remain "logically sensitive" to such criticism.
4. an equal distribution of intellectual authority "among qualified practitioners" (Longino, 1990, 76). Rather like Merton's norm of universalism, this criterion prohibits the scientific community from privileging the arguments of one investigator over another on the basis of attributes such

[9] The preeminence of the scientific journal was more secure when Longino wrote her 1990 book than it is today, when scientists post pre-publication drafts and other documents to pre-print archives such as arXiv (arXiv.org).

as gender, race, or religious affiliation. Longino notes some historical and ongoing threats to or outright violations of this criterion, such as the "bureaucratization of science" and the "exclusion, whether overt or more subtle, of women and members of certain racial minorities from scientific education and the scientific professions" (Longino, 1990, 78).

12.5 Feminist critiques

Lurking in the background of our brief discussion of explanatory models in the study of gender and behavior is the way that values and norms regarding gender have underpinned dominant assumptions. *Feminist* critical studies of science (Wylie, Potter, & Bauchspies, 2012), which seek to scrutinize the roles that gender and gender norms have played and continue to play in science, constitute an important part of the debates over how values figure in scientific inquiry. This literature, which also contributes to *feminist epistemology* (E. Anderson, 2012), has grown quite large, such that even a brief characterization of the chief positions and arguments would require a separate chapter. For the present, I will mention just two points of contrast to indicate the relevance of Longino's account, and of feminist critiques more generally, to these debates.

Numerous historical studies have shown how different gender-related values have influenced scientific theorizing. Longino herself discusses, for example, the very direct roles that *androcentrism* (the "perception of social life from a male point of view"), *sexism* ("statements, attitudes, practices, behavior, or theories presupposing, asserting, or implying the inferiority of women, the legitimacy of their subordination, or the legitimacy of sex-based prescriptions of social roles and behaviors"), and *heterosexism* (in part, "the idea that 'they' are made for and hence complementary to 'us' ") have played in the study of human evolution (Longino, 1990, 128–132). Numerous studies have documented such biases in various aspects of the field of human evolutionary studies (e.g., Dahlberg, 1981; Lloyd, 2005). What is the remedy for this?

12.5.1 Longino's pluralism in contrast with value-neutrality and standpoint theory

Some critics (including some feminists) have responded to problems of androcentric and other biases with a call for a return to the ideal of value-free science that prohibits scientists from allowing values to influence their

interpretations of their data and their conclusions from those data. According to this line of thought, the best philosophical response to the problem of gender bias in scientific theorizing (and the best strategy for achieving feminist goals) is the removal of all bias from science (Pinnick, Koertge, & Almeder, 2003).

For example, critics have noted the androcentric character of the *man-the-hunter* approach to the interpretation of archaeological evidence regarding early hominids. Researchers operating within this framework assumed that tools were developed for the purposes of hunting and that the hunting activities of males drove the process of human evolution. In response, feminist archaeologists developed the *woman-the-gatherer* framework, regarding the activities of women as the main evolutionary force (Dahlberg, 1981). From the standpoint of the advocates of value-free science, both frameworks would seem to constitute failures of objectivity, in which researchers have allowed their personal biases to influence their handling of data.

Longino's response acknowledges that the woman-the-gatherer framework "is as gynocentric as its rival is androcentric," but notes that "its great value from a logical point of view is its revelation of the epistemologically arbitrary character of the man-the-hunter framework. As long as both frameworks offer coherent and comprehensive accounts of the relevant data, neither can displace the other" (Longino, 1990, 130). According to Longino's analysis of evidential argumentation, the data offered by archaeological evidence cannot be used to draw conclusions about human evolution without invoking some kind of explanatory model, and the needed models themselves lack sufficient evidential backing to remove a degree of arbitrariness in the choice between them. This, for Longino, makes the value-free approach unavailable. A proliferation of distinct value-inflected approaches enables us to draw explicit contrasts and critically examine the values on which these theories depend. In this way, Longino's approach resembles Feyerabend's advocacy of theory proliferation. Whereas Feyerabend's pluralism was part of his thoroughgoing skepticism about the value of scientific ideals like objectivity, however, Longino eschews epistemological anarchism in favor of a contextualist and pluralist account of objectivity.

Whereas the value-free approach seeks to eliminate masculinist values by driving out all values from the interpretation of data and the drawing of conclusions from data, and Longino's pluralism calls for a proliferation of value-laden theoretical approaches, *feminist standpoint epistemology* seeks to

replace (or at least demote) masculinist values with feminist values. More carefully, standpoint theory rests on the premise that the different social perspectives that individuals occupy result in their having different abilities to gain cognitive access to certain kinds of facts. Marxist standpoint theory, for example, holds that members of the proletariat, by organizing and becoming conscious of their role in the capitalist system, have "epistemic privilege over fundamental questions of economics, sociology, and history" (E. Anderson, 2012). Nancy Hartsock proposes that feminists appropriate this Marxian approach: because "women's lives differ structurally from those of men" they "make available a particular and privileged vantage point on male supremacy . . . which can ground a powerful critique of the phallocratic institutions and ideology which constitute the capitalist form of patriarchy" (Hartsock, 1983, 36).

In a less explicitly Marxian vein, Sandra Harding argues that standpoint theory allows for the articulation of a concept of *strong objectivity* that is superior to the conception of objectivity in terms of *neutrality*. The latter, according to Harding and other standpoint theorists, is an illusion: "It is a delusion – and a historically identifiable one – to think that human thought could completely erase the fingerprints that reveal its production process." That process, she argues, while it succeeds in weeding out the influence of interests that vary amongst different researchers, leaves intact the interests that are "shared by all (or virtually all) of the observers," so that "culturewide assumptions *that have not been criticized within the scientific research process* are transported into the results of research" (Harding, 1993, 57; emphasis in the original). Whereas Longino's response to the elusiveness of neutrality is to advocate a pluralistic conception of objectivity in terms of the proliferation of values within a community organized in accordance with the criteria enumerated above, Harding's response is to valorize the perspective of marginalized people with regard at least to certain kinds of phenomena: "Democracy-advancing values have systematically generated less partial and distorted beliefs than others" (Harding, 1993, 71).

Whether Harding is correct here is, as Cassandra Pinnick points out, an empirical question. (Pinnick finds Harding's position deficient in empirical support; Pinnick, 1994). Longino objects to theories that "recommend an alternate subject position as epistemically superior" on the grounds that "arguments are missing" and in principle must remain so. Empirical arguments for claims of superior epistemic privilege would require first

identifying the standpoint from which one can make an adequate appraisal of the evidence. A priori arguments for the same would have to both claim "a particular structure for the world" and identify "a particular subjectivity as uniquely capable of knowing that structure." Either way the resulting argument is bound to beg the question (Longino, 1993, 106–110). Noting that "it's not clear that any particular subject position could be adequate to generate knowledge," Longino locates objectivity not in the epistemic advantage of particular standpoints over others, but in the socially organized interaction of scientists occupying different perspectives.[10] Meanwhile, Alison Wylie and Lynn Hankinson Nelson have embarked upon an empirical assessment of feminist interventions in science with a discussion of case studies (Wylie & Hankinson Nelson, 2007).

12.6 Conclusion: underdetermination and honesty

The ideal of a value-free science faces challenges to its achievability as well as to its usefulness or even coherence as an ideal. Douglas has argued persuasively that, taken in too strong a sense, this ideal is incompatible with an acknowledgment of the moral responsibilities of scientists (Douglas, 2009). But just where and how can values enter into scientific reasoning without compromising the objectivity of scientific inquiry?

Longino's argument that values enter into the use of evidence in support of scientific theories relies on an appeal to a version of the underdetermination thesis: treating data as evidence relevant to a given hypothesis requires the use of an explanatory model, which may lack independent warrant. Contextual values, for Longino, contribute to a "rich pool of varied resources, constraints, and incentives to help close the gap left by logic" (Longino, 2002, 128).

As Kristen Intemann has argued, this appeal to an underdetermination gap does not suffice to show that contextual values have a legitimate role to play in determining evidential relationships between data and hypotheses (Intemann, 2005). After all, where such gaps do arise, it may be possible to fill them by relying on purely descriptive claims, and where that is not possible, it may be that the epistemically responsible choice would be to suspend

[10] Kathleen Okruhlik also advocates a pluralist approach to feminist philosophy of science in her (1994).

judgment regarding the hypotheses at issue. Moreover, Deborah Mayo, operating within an error-statistical framework, has argued that assumptions needed to close the logical gap between data and hypothesis can often be provided with their own evidential warrant (Mayo, 1996, ch. 6; Mayo, 1997a).

These comments do not, however, lead directly to a vindication of the value-free ideal of science, since, as also argued by Intemann, contextual values may nonetheless play an important role to the extent that scientific hypotheses themselves have content that implies or supports certain contextual values (in opposition to what Lacey calls the thesis of *neutrality*) or to the extent that the aims of science implicate certain contextual values. The former claim has been defended by numerous scholars (e.g., E. Anderson, 1995; Fausto-Sterling, 2000). Philip Kitcher defends the latter claim in his (2001). Moreover, the worry that allowing value judgments into evidential reasoning will leave conclusions dependent on arbitrary or unreasoned judgments assumes that value judgments themselves cannot be provided with evidential warrant. One might, as Elizabeth Anderson argues, reject that assumption, opening up the possibility that value judgments might lend epistemic support to conclusions drawn from scientific data in virtue of enjoying independent support of their own (E. Anderson, 2004).

The value-free ideal, as articulated in Lacey's impartiality thesis especially, enjoys a strong intuitive appeal. "Only the data matter," scientists will sometimes claim. We might imagine that data are unambiguous, quantitative, and not up for debate, while values are fuzzy, qualitative, and subject to disagreement. By now it should be evident that this contrast at best constitutes a drastic overstatement and at worst is dangerously misleading. Longino is surely correct (and in agreement with most other philosophers of science) in denying that data display their evidential significance *self-evidently*. Whether we describe it in terms of models, assumptions, or auxiliary hypotheses, data need to inhabit some richer structure in order to become evidentially *useful*. The important questions regarding the possibilities of impartiality and neutrality in science concern the roles of value judgments in undergirding such structures and the epistemic status of such judgments.

To the extent that values play an *unavoidable* role in some areas of science, as Longino might argue for the study of human behavior, the value-free ideal of science poses a threat to science. If the conception of 'good science' requires that one exclude contextual values from playing a role in the conclusions one draws from data, yet it is impossible in a given field to meet that requirement,

then we may expect investigators to seek public acceptance of their work by obscuring the role that values play. Better, under such circumstances, for scientists to feel that they are free to discuss openly the ways in which values influence their interpretations and inferences from the data and the kind of reasoning that leads them to rely on those values and not others.

When scientists make explicit the supporting structures that they use to turn data into evidence, they can then turn to the crucial task of asking how those structures might fail to be adequate for the inferential task at hand.[11] I have elsewhere referred to this task as the *securing* of evidence. To secure an evidential inference from a given body of data is, roughly, to determine the various ways the world might be, given the relevant state of knowledge, such that the data in question would fail to be good evidence for the conclusion drawn (Staley & Cobb, 2011; Staley, 2012).[12]

Attending to the security of evidence requires an explicit consideration of the kinds of knowledge one brings to bear on a given body of data, as well as a comprehensive consideration of the possibilities for error at each stage of the scientific process. In this way it constitutes an extension of the error-statistical orientation discussed in Chapter 9: rather than asking how probable or well confirmed a scientific claim is, we might ask instead about the ways in which it might be in error, and what steps investigators have taken to eliminate or otherwise take into account such errors. If we learned that scientists reached a conclusion in part by relying on value-laden assumptions on no stronger basis than that those assumptions reinforce their own subjective preferences, then we should certainly consider the inference to suffer from a significant potential source of error that had not been secured. On the other hand, should we find that reliance on a contextual value in a

[11] I state the target in terms of adequacy rather than truth because sometimes the auxiliary assumptions that scientists rely upon are sufficient for this task without being true, perhaps because a simplified model is sufficient for the level of accuracy that one desires, or perhaps because one makes overly 'conservative' assumptions that one knows to be false but that reduce the risk of committing some particularly undesirable error.

[12] Nancy Cartwright has likewise called for making assumptions explicit that are necessary for extrapolating from the conditions under which data has been collected to contexts in which conclusions from those data are to be applied. She has particularly called attention to this issue in the context of "evidence-based policy making" where the dangers of mistakenly inferring from the effectiveness of a policy intervention under one set of conditions to its effectiveness under other circumstances are especially grave (Cartwright, 2012).

given case was warranted with further evidence, or that it promoted the aim for which the inference was drawn, this could contribute to the securing of the inference at hand.

We began this book by noting the tremendous impact and authority of scientists and the work they do. That great responsibility comes along with this power is a truism, but we have in a sense been discussing through all of these chapters the nature of that responsibility. What is the undertaking of science? To what standards should we hold scientists beyond the norms that we consider applicable to all human behavior generally? One point that I hope has emerged in this chapter is that science cannot be regarded as an entirely detached, disinterested enterprise. To undertake scientific inquiry itself reflects a certain kind of commitment and a certain judgment about what is worth doing. Deciding what kind of scientific inquiry to conduct and how to carry it out requires further evaluative deliberation. Whatever further roles contextual values might play, the securing of evidence and the susceptibility of conclusions to error-probing and transformative criticism demand that those roles be laid bare. For such an undertaking, a partnership of philosophers and scientists may prove fruitful, or even essential.

References

Aad, G., Abajyan, T., Abbott, J., et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters*, *B716*, 1–29.

Achinstein, P. (1968). *Concepts of science: A philosophical analysis*. Baltimore, MD: The Johns Hopkins Press.

(1983). *The nature of explanation*. New York: Oxford University Press.

(2000). Proliferation: Is it a good thing? In J. Preston, G. Munévar, & D. Lamb (eds.), *The worst enemy of science? Essays in memory of Paul Feyerabend* (pp. 37–46). New York: Oxford University Press.

(2001). *The book of evidence*. New York: Oxford University Press.

(2002). Is there a valid experimental argument for scientific realism? *The Journal of Philosophy*, *99*, 470–495.

(2010). Mill's sins or Mayo's errors? In D. G. Mayo & A. Spanos (eds.), *Error and inference: Recent exchanges on experimental reasoning, reliability, objectivity, and rationality* (pp. 170–188). New York: Cambridge University Press.

Ackermann, R. (1985). *Data, instruments, and theory*. Princeton University Press.

(1989). The new experimentalism. *The British Journal for the Philosophy of Science*, *40*(2), 185–190.

Anderson, C. (1933). The positive electron. *Physical Review*, *43*, 491–494.

(1999). *The discovery of anti-matter: The autobiography of Carl David Anderson, the youngest man to win the Nobel prize* (R. J. Weiss, ed.). Singapore: World Scientific.

Anderson, E. (1995). Knowledge, human interests, and objectivity in feminist epistemology. *Philosophical Topics*, *23*, 27–57.

(2004). Uses of value judgments in science: A general argument, with lessons from a case study of feminist research on divorce. *Hypatia*, *19*(1), 1–24.

(2012). Feminist epistemology and philosophy of science. In E. N. Zalta (ed.), *The Stanford encyclopedia of philosophy* (Fall 2012 ed.). http://plato.stanford.edu/archives/fall2012/entries/feminism-epistemology/.

Armstrong, D. M. (1983). *What is a law of nature?* Cambridge University Press.

Ashford, N., Ryan, C. W., & Caldart, C. C. (1983). A hard look at federal regulation of formaldehyde: A departure from reasoned decisionmaking. *Harvard Environmental Law Review*, *7*, 297–370.

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Ball, L., Mankiw, N. G., & Nordhaus, W. D. (1999). Aggregate demand and long-run unemployment. *Brookings Papers on Economic Activity*, *1999*(2), 189–251.

Barker-Plummer, D., Barwise, J., & Etchemendy, J. (2011). *Language, proof, and logic* (Second edn.). Stanford, CA: Center for the Study of Language and Information.

Barnes, E. (1992). Explanatory unification and the problem of asymmetry. *Philosophy of Science*, *59*(4), 558–571.

Batterman, R. W. (2001). *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence*. New York: Oxford University Press.

Baum, R., & Sheehan, W. (2003). *In search of planet Vulcan: The ghost in Newton's clockwork universe*. New York: Basic Books.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418.

Bekelman, J. E., Li, Y., & Gross, C. P. (2003). Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *JAMA*, *289*(4), 454–465.

Berger, J., & Bernardo, J. (1992). On the development of reference priors. In J. M. Bernardo, J. O. Berger, P. Dawid, & A. Smith (eds.), *Bayesian statistics 4* (pp. 35–60). New York: Oxford University Press.

Bernardini, C. (2004). AdA: The first electron-positron collider. *Physics in Perspective*, *6*, 156–183.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society*, *B41*, 113–147.

Bhattacharjee, Y. (2013, April 26). The mind of a con man. *New York Times*.

Bird, A. (2000). *Thomas Kuhn*. Princeton University Press.

Birnbaum, A. (1977). The Neyman-Pearson theory as decision theory, and as inference theory; with a criticism of the Lindley-Savage argument for Bayesian theory. *Synthese*, *36*(1), 19–49.

Blackett, P., & Occhialini, G. (1933). Some photographs of the tracks of penetrating radiation. *Proceedings of the Royal Society of London*, *139A*, 699–726.

Blackwell, R. J. (2006). *Behind the scenes at Galileo's trial*. Notre Dame, IN: University of Notre Dame Press.

Blanchard, O., & Jimeno, J. F. (1995). Structural unemployment: Spain versus Portugal. *The American Economic Review*, *85*(2), 212–218.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, *97*(3), 303–352.

Bokulich, A. (2008). *Reexamining the quantum–classical relation: Beyond reductionism and pluralism.* Cambridge Universiy Press.

Born, M., Heisenberg, W., & Jordan, P. (1926). Zur Quantenmechanik ii. *Zeitschrift für Physik*, *35*, 557–615.

Born, M., & Jordan, P. (1925). Zur Quantenmechanik. *Zeitschrift für Physik*, *34*, 858–888.

Boyd, R. (1989). What realism implies and what it does not. *Dialectica*, *43*, 5–29.

Braithwaite, R. B. (1953). *Scientific explanation: A study of the function of theory, probability and law in science*. Cambridge University Press.

Bromberger, S. (1966). Why-questions. In B. Brody (ed.), *Readings in the philosophy of science* (pp. 66–84). Englewood Cliffs, NJ: Prentice Hall.

Carnap, R. (1936). Testability and meaning. *Philosophy of Science*, *3*(4), 419–471.

(1937). Testability and meaning–continued. *Philosophy of Science*, *4*(1), 1–40.

(1955). Foundations of logic and mathematics. In O. Neurath, R. Carnap, & C. W. Morris (eds.), *International encyclopedia of unified science* (Vol. 1, pp. 139–213). University of Chicago Press.

(1962). *Logical foundations of probability*. University of Chicago Press.

(1963). Replies and systematic expositions. In P. A. Schilpp (ed.), *The philosophy of Rudolf Carnap* (pp. 859–1013). LaSalle, IL: Open Court.

(1968). Inductive logic and inductive intuition. In I. Lakatos (ed.), *The problem of inductive logic* (pp. 258–267). Amsterdam: North Holland.

Cartwright, N. (1983). *How the laws of physics lie*. New York: Oxford University Press.

(2012). Will this policy work for you? Predicting effectiveness better: How philosophy helps. *Philosophy of Science*, *79*(5), 973–989.

Chaloner, C. (1997). The most wonderful experiment in the world: A history of the cloud chamber. *The British Journal for the History of Science*, *30*(3), 357–374.

Chatrchyan, S., Barass, T. A., Bostock, F. J. D., et al. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters*, *B716*, 30–61.

Churchman, C. W. (1948). Statistics, pragmatics, induction. *Philosophy of Science*, *15*(3), 249–268.

Cobb, A. (2009). Michael Faraday's 'historical sketch of electro-magnetism' and the theory-dependence of experimentation. *Philosophy of Science*, *76*, 624–636.

Committee on the Use of Chimpanzees in Biomedical and Behavioral Research. (2011). *Chimpanzees in biomedical and behavioral research: Assessing the necessity* (B. M. Altevogt, D. E. Pankevich, M. K. Shelton-Davenport, & J. P. Kahn, eds.). Washington, DC: The National Academies Press.

Cox, D. R. (2006). *Principles of statistical inference*. New York: Cambridge University Press.

Cox, D. R., & Hinkley, D. V. (1974). *Theoretical statistics*. London: Chapman and Hall.

Cox, D. R., & Mayo, D. G. (2010). Objectivity and conditionality in frequentist inference. In D. G. Mayo & A. Spanos (eds.), *Error and inference: Recent exchanges on experimental reasoning, reliability, objectivity, and rationality* (pp. 276–304). New York: Cambridge University Press.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press.

(2008). Physical law and mechanistic explanation in the Hodgkin and Huxley model of the action potential. *Philosophy of Science*, *75*(5), 1022–1033.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.

Dahlberg, F. (ed.). (1981). *Woman the gatherer*. New Haven: Yale University Press.

Darwin, C. (1859). *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life*. London: John Murray.

Davidson, J. (2013, January 17). Federal scientists can again research gun violence. *Washington Post*.

de Regt, H. W. (2004). Review of James Woodward, *Making Things Happen*. *Notre Dame Philosophical Reviews*. Available from http://ndpr.nd.edu/news/23818-making-things-happen-a-theory-of-causal-explanation/

De Finetti, B. ([1937]1964). Foresight: Its logical laws, its subjective sources (H. E. Kyburg, trans.). In H. E. Kyburg & H. E. Smokler (eds.), *Studies in subjective probability* (pp. 93–158). New York: John Wiley and Sons.

(1972). *Probability, induction, and statistics: The art of guessing*. New York: John Wiley.

Dirac, P. A. M. (1928). The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *117*(778), 610–624.

(1931). Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society of London*, *133A*, 60–72.

Doob, J. L. (1971). What is a martingale? *American Mathematical Monthly*, *78*, 451–462.

Dorling, J. (1979). Bayesian personalism, the methodology of scientific research programmes, and Duhem's problem. *Studies in History and Philosophy of Science*, *10*, 177–187.

Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press.

Dowe, P. (1992). Wesley Salmon's process theory of causality and the conserved quantity theory. *Philosophy of Science*, *59*(2), 195–216.

(1995). Causality and conserved quantities: A reply to Salmon. *Philosophy of Science*, *62*(2), 321–333.

Duhem, P. (1991). *The aim and structure of physical theory*. Princeton University Press.

Earman, J. (1986). *A primer on determinism*. Dordrecht: Reidel.

(1992). *Bayes or bust? a critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.

Einstein, A. (1918). Prinzipielles zur allgemeinen Relativitätstheorie. *Annalen der Physik*, *55*, 241–244.

Elliott, K. C. (2011). Direct and indirect roles for values in science. *Philosophy of Science*, *78*, 303–324.

Enderton, H. (2001). *A mathematical introduction to logic* (Second edn.). New York: Academic Press.

Enz, C. P. (2002). *No time to be brief: A scientific biography of Wolfgang Pauli*. New York: Oxford University Press.

Epstein, S. S. (1998). *The politics of cancer revisited*. Fremont Center, NY: East Ridge.

Fausto-Sterling, A. (2000). *Sexing the body: Gender politics and the construction of sexuality*. New York: Basic Books.

Feyerabend, P. (1970). Against method: Outline of an anarchistic theory of knowledge. In *Analyses of theories and methods of physics and psychology* (Vol. IV, pp. 17–130). Minneapolis, MN: University of Minnesota Press.

(1978). *Science in a free society*. London: Verso.

(1981). Two models of epistemic change: Mill and Hegel. In *Problems of empiricism.* Cambridge University Press.

(1988). *Against method* (Revised edn.). London: Verso.

(1991). *Three dialogues on knowledge*. Oxford: Basil Blackwell.

(1995). *Killing time: The autobiography of Paul Feyerabend*. University of Chicago Press.

Fine, A. (1984). The natural ontological attitude. In J. Leplin (Ed.), *Scientific realism* (pp. 83–107). Berkeley, CA: University of California Press.

(1986). Unnatural attitudes: Realist and instrumentalist attachments to science. *Mind*, *95*, 149–179.

Finocchiaro, M. (1989). *The Galileo affair: A documentary history*. Berkeley, CA: University of California Press.

Fiorin, G., Carnevale, V., & DeGrado, W. F. (2010). The flu's proton escort. *Science*, *330*(6003), 456–458.

Fisher, R. A. (1949). *The design of experiments* (Fifth edn.). New York: Hafner Publishing Co.

Franklin, A. (1986). *The neglect of experiment*. New York: Cambridge University Press.

(2005). *No easy answers: Science and the pursuit of knowledge*. University of Pittsburgh Press.

(2012). Experiment in physics. In E. N. Zalta (ed.), *The Stanford encyclopedia of philosophy* (Winter 2012 ed.). http://plato.stanford.edu/archives/win2012/entries/physics-experiment/.

Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy*, *71*(1), 5–19.

(1999). *Reconsidering logical positivism*. Cambridge University Press.

Fuller, S. (2001). *Thomas Kuhn: A philosophical history for our times*. University of Chicago Press.

Gaifman, H., & Snir, M. (1982). Probabilities over rich languages. *Journal of Symbolic Logic*, *47*, 495–548.

Galilei, G. ([1632]1967). *Dialogue concerning the two chief world systems* (Revised edn.; S. Drake, trans.). Berkeley, CA: University of California Press.

Galison, P. (1987). *How experiments end*. University of Chicago Press.

Galison, P., & Assmus, A. (1989). Artificial clouds, real particles. In D. Gooding, T. Pinch, & S. Schaffer (eds.), *The uses of experiment: Studies in the natural sciences* (pp. 225–273). New York: Cambridge University Press.

Giere, R. N. (1999). *Science without laws*. University of Chicago Press.

Gillies, D. (1993). The Duhem thesis and the Quine thesis. In M. Curd & J. A. Cover (eds.), *Philosophy of science: The central issues* (pp. 302–319). New York: W. W. Norton.

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*(1), 49–71.

Gooding, D. (1990). *Experiment and the making of meaning: Human agency in scientific observation and experiment*. Dordrecht: Kluwer.

Gorman, J. (2011, December 15). U.S. will not finance new research on chimps. *New York Times*.

(2013, January 22). Agency moves to retire most research chimps. *New York Times*.

Grene, M. (1966). *The knower and the known*. New York: Basic Books.

(1985). Perception, interpretation, and the sciences. In D. Depew & B. Weber (eds.), *Evolution at a crossroads* (pp. 1–20). Cambridge, MA: MIT Press.

Guerlac, H. (1983). Can we date Newton's early optical experiments? *Isis*, *74*, 74–80.

Hacking, I. (1975). *The emergence of probability*. Cambridge University Press.

(1981). Do we see through a microscope? *Pacific Philosophical Quarterly*, *62*, 305–322.

(1982). Experimentation and scientific realism. *Philosophical Topics*, *13*, 71–87.

(1983). *Representing and intervening*. Cambridge University Press.

(1990). *The taming of chance*. Cambridge University Press.

(2001). *An introduction to probability and inductive logic*. Cambridge University Press.

Hájek, A. (2008). Arguments for – or against – probabilism? *The British Journal for the Philosophy of Science*, *59*, 793–819.

Hanson, N. R. (1958). *Patterns of discovery*. Cambridge University Press.

(1962). Leverrier: The zenith and nadir of Newtonian mechanics. *Isis*, *53*(3), 359–378.

Harding, S. (1993). Rethinking standpoint epistemology: What is 'strong objectivity'? In L. Alcoff & E. Potter (eds.), *Feminist epistemologies* (pp. 49–82). New York: Routledge. (Rpt. in Harding 2004.)

(ed.). (2004). *The feminist standpoint theory reader*. New York: Routledge.

Harman, G. (1965). The inference to the best explanation. *The Philosophical Review*, *74*, 88–95.

Harper, W. (2002). Newton's argument for universal gravitation. In I. B. Cohen & G. E. Smith (eds.), *The Cambridge companion to Newton* (pp. 174–201). New York: Cambridge University Press.

(2007). Newton's methodology and Mercury's perihelion before and after Einstein. *Philosophy of Science*, *74*(5), 932–942.

(2011). *Isaac Newton's scientific method: Turning data into evidence about gravity and cosmology*. New York: Oxford University Press.

Hartsock, N. C. M. (1983). The feminist standpoint: Developing the ground for a specifically feminist historical materialism. In M. B. Hintikka & S. Harding (eds.), *Discovering reality* (pp. 283–310). Boston, MA: D. Reidel. (Rpt. in Harding 2004.)

Heisenberg, W. (1925). Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift für Physik*, *33*, 261–276.

Hempel, C. G. (1965). Empiricist criteria of cognitive significance: Problems and changes. In *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 101–119). New York: The Free Press.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*, 135–175.

Hesse, M. (1980). *Revolutions and reconstructions in the philosophy of science*. Bloomington, IN: Indiana University Press.

Hillis, D. M., Moritz, C., & Mable, B. K. (eds.). (1996). *Molecular systematics* (Second edn.). Sunderland, MA: Sinauer Associates.

Hitchcock, C. R. (1995). Salmon on explanatory relevance. *Philosophy of Science*, *62*(2), 304–320.

Hodgkin, A., & Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, *117*, 500–544.

Hon, G. (1987). H. Hertz: 'The electrostatic and electromagnetic properties of the cathode rays are either nil or very feeble.' (1883) A case-study of an experimental error. *Studies In History and Philosophy of Science Part A*, *18*(3), 367–382.

(1989). Towards a typology of experimental errors: An epistemological view. *Studies in the History and Philosophy of Science*, *20*, 469–504.

Hooper, W. (1998). Inertial problems in Galileo's preinertial framework. In P. Machamer (ed.), *The Cambridge companion to Galileo* (pp. 146–174). New York: Cambridge University Press.

Howson, C. (1997). Error probabilities in error. *Philosophy of Science*, *64*, S185–S194.

Howson, C., & Urbach, P. (2005). *Scientific reasoning: The Bayesian approach* (Third edn.). Chicago: Open Court.

Hoyningen-Huene, P. (1993). *Reconstructing scientific revolutions: Thomas S. Kuhn's philosophy of science*. University of Chicago Press.

(2000). Paul K. Feyerabend: An obituary. In J. Preston, G. Munévar, & D. Lamb (eds.), *The worst enemy of science? Essays in memory of Paul Feyerabend* (pp. 3–15). New York: Oxford University Press.

Hume, D. ([1748]1993). *An enquiry concerning human understanding* (Second edn.; E. Steinberg, ed.). Indianapolis: Hackett Publishing Company.

Humphreys, P. (2008). *Notes on the origin of the term 'Dutch book'*. Retrieved July 20, 2012, from http://people.virginia.edu/~pwh2a/dutch%2520book%2520origins.doc

Intemann, K. (2005). Feminism, underdetermination, and values in science. *Philosophy of Science*, *72*(5), 1001–1012.

Jasanoff, S. S. (1987). Contested boundaries in policy-relevant science. *Social Studies of Science*, *17*(2), 195–230.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. New York: Cambridge University Press.

Jeffrey, R. C. (1956). Valuation and acceptance of scientific hypotheses. *Philosophy of Science*, *23*(3), 237–246.

Kadvany, J. (2001). *Imre Lakatos and the guises of reason*. Durham, NC: Duke University Press.

Kellermann, A. L., & Rivara, F. P. (2013). Silencing the science on gun research. *JAMA*, *309*(6), 549–550.

Kerns, W. D., Pavkov, K. L., Donofrio, D. J., Gralla, E. J., & Swenberg, J. A. (1983). Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. *Cancer Research*, *43*(9), 4382–4392.

Kitcher, P. (1976). Explanation, conjunction, and unification. *The Journal of Philosophy*, *73*(8), 207–212.

(1981). Explanatory unification. *Philosophy of Science*, *48*(4), 507–531.

(1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (eds.), *Scientific explanation* (Vol. 13, pp. 410–506). Minneapolis, MN: University of Minnesota Press.

(2001). *Science, truth, and democracy*. New York: Oxford University Press.

Kolmogorov, A. N. (1950). *Foundations of the theory of probability* (N. Morrison, trans.). New York: Chelsea Pub. Co.

Kourany, J. (2010). *Philosophy of science after feminism*. New York: Oxford University Press.

Kronig, R. (1960). The turning point. In M. Fierz & V. F. Weisskopf (eds.), *Theoretical physics in the twentieth century: A memorial volume to Wolfgang Pauli* (pp. 5–39). New York: Interscience Publishers.

Kuhn, T. (1957). *The Copernican revolution: Planetary astronomy in the development of western thought*. Cambridge, MA: Harvard University Press.

(1962). *The structure of scientific revolutions*. University of Chicago Press.

(1977). Objectivity, value judgment, and theory choice. In *The essential tension* (pp. 320–339). University of Chicago Press.

(1993). Afterwords. In P. Horwich (ed.), *World changes: Thomas Kuhn and the nature of science* (pp. 311–341). Cambridge, MA: The MIT Press.

(1996). *The structure of scientific revolutions, third edition*. University of Chicago Press.

Kyburg, H. E. (1965). Salmon's paper. *Philosophy of Science*, *32*(2), 147–151.

(1993). Peirce and statistics. In E. C. Moore (ed.), *Charles S. Peirce and the philosophy of science: Papers from the Harvard sesquicentennial congress* (pp. 130–138). Tuscaloosa, AL: University of Alabama Press.

(1978). Subjective probability: Criticisms, reflections, and problems. *Journal of Philosophical Logic*, *7*(1), 157–180.

Lacey, H. (1999). *Is science value free? values and scientific understanding*. London: Routledge.

(2005). *Values and objectivity in science: The current controversy about transgenic crops*. Oxford: Lexington Books.

Ladyman, J., Ross, D., Spurrett, D., & Collier, J. (2007). *Every thing must go: Metaphysics naturalized*. New York: Oxford University Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge University Press.

(1978). *The methodology of scientific research programmes: Philosophical papers* (Vol. 1; J. Worrall & G. Currie, eds.). Cambridge University Press.

Lakatos, I., & Feyerabend, P. (1999). *For and against method: Including Lakatos's lectures on scientific method and the Lakatos-Feyerabend correspondence* (M. Motterlini, ed.). University of Chicago Press.

Lakatos, I., & Musgrave, A. (eds.). (1970). *Criticism and the growth of knowledge*. Cambridge University Press.

Lange, M. (2009). *Laws and lawmakers: Science, metaphysics, and the laws of nature*. New York: Oxford University Press.

Larvor, B. (1998). *Lakatos: An introduction*. New York: Routledge.

Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science*, *48*, 19–49.

(1996). Demystifying underdetermination. In *Beyond positivism and relativism* (pp. 29–54). Boulder, CO: Westview Press.

Lemmens, T., & Waring, D. R. (eds.). (2006). *Law and ethics in biomedical research: Regulation, conflict of interest, and liability*. University of Toronto Press.

Leplin, J. (1997). *A novel defense of scientific realism*. New York: Oxford University Press.

Levi, I. (1962). On the seriousness of mistakes. *Philosophy of Science*, *29*(1), 47–65.

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.

(1986). *Philosophical papers: Volume II*. New York: Oxford University Press.

Lewis, P. J. (2001). Why the pessimistic induction is a fallacy. *Synthese*, *129*, 371–380.

Lilford, R. J., & Jackson, J. (1995). Equipoise and the ethics of randomization. *Journal of the Royal Society of Medicine*, *88*, 552–559.

Linnen, C. R., & Farrell, B. D. (2008). Comparison of methods for species-tree inference in the sawfly genus *Neodiprion* (hymenoptera: Diprionidae). *Systematic Biology*, *57*(6), 876–890.

Lipton, P. (2004). *Inference to the best explanation* (Second edn.). London: Routledge.

Lloyd, E. (1997). Feyerabend, Mill, and pluralism. *Philosophy of Science*, *64*, S396–S407.

(2005). *The case of the female orgasm: Bias in the science of evolution*. Cambridge, MA: Harvard University Press.

Longino, H. (1990). *Science as social knowledge*. Princeton University Press.

(1993). Subjects, power, and knowledge: Description and prescription in feminist philosophies of science. In L. Alcoff & E. Potter (eds.), *Feminist epistemologies* (pp. 101–120). New York: Routledge.

(1994). The fate of knowledge in social theories of science. In F. Schmitt (ed.), *Socializing epistemology: The social dimensions of knowledge* (pp. 135–157). Lanham, MD: Rowman and Littlefield Publishers, Inc.

(2002). *The fate of knowledge*. Princeton University Press.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25.

MacIntyre, A. (1981). *After virtue*. Notre Dame, IN: University of Notre Dame Press.

Magnus, P. D., & Callender, C. (2004). Realist ennui and the base rate fallacy. *Philosophy of Science*, *71*(3), 320–338.

Massimi, M. (2004). Non-defensible middle ground for experimental realism: Why we are justified to believe in colored quarks. *Philosophy of Science*, *71*, 36–60.

Masterman, M. (1970). The nature of a paradigm. In I. Lakatos & A. Musgrave (eds.), *Criticism and the growth of knowledge* (pp. 59–89). Cambridge University Press.

Maxwell, G. (1962). The ontological status of theoretical entities. In H. Feigl & G. Maxwell (eds.), *Minnesota studies in the philosophy of science* (Vol. 3: Scientific Explanation, Space, and Time, pp. 3–27). Minneapolis, MN: University of Minnesota Press.

Mayo, D. G. (1985). Behavioristic, evidentialist, and learning models of statistical testing. *Philosophy of Science*, *52*, 493–516.

(1991). Sociological versus metascientific views of risk assessment. In D. G. Mayo & R. Hollander (eds.), *Acceptable evidence: Science and values in risk management* (pp. 249–279). New York: Oxford University Press.

(1993). The test of experiment: C. S. Peirce and E. S. Pearson. In E. C. Moore (ed.), *Charles S. Peirce and the philosophy of science: Papers from the Harvard sesquicentennial congress* (pp. 161–174). Tuscaloosa, AL: University of Alabama Press.

(1996). *Error and the growth of experimental knowledge*. University of Chicago Press.

(1997a). Duhem's problem, the Bayesian way, and error statistics, or 'what's belief got to do with it?' *Philosophy of Science*, *64*, 222–244.

(1997b). Response to Howson and Laudan. *Philosophy of Science*, *64*(2), 323–333.

(2002). Theory testing, statistical methodology, and the growth of experimental knowledge. In P. Gärdenfors, J. Wolinski, & K. Kijania-Placek (eds.), *In the scope of logic, methodology, and philosophy of science* (pp. 171–190). Dordrecht: Kluwer.

Mayo, D. G., & Spanos, A. (2004). Methodology in practice: Statistical misspecification testing. *Philosophy of Science*, *71*, 1007–1025.

(2006). Severe testing as a basic concept in a Neyman-Pearson philosophy of induction. *The British Journal for the Philosophy of Science*, *57*(2), 323–357.

(eds.). (2009). *Error and inference: Recent exchanges on experimental reasoning, reliability, objectivity, and rationality*. New York: Cambridge University Press.

McAllister, J. A. (1999). *Beauty and revolution in science*. Ithaca, NY: Cornell University Press.

McMullin, E. (1982). Values in science. In P. D. Asquith & T. Nickles (eds.), *PSA 1982: Proceedings of the biennial meeting of the Philosophy of Science Association* (Vol. II: Symposia and Invited Papers, pp. 3–28). University of Chicago Press.

(1984). A case for scientific realism. In J. Leplin (ed.), *Scientific realism* (pp. 8–40). Berkeley, CA: University of California Press.

(1993). Rationality and paradigm change in science. In P. Horwich (ed.), *World changes* (pp. 55–78). Cambridge, MA: MIT Press.

Merton, R. K. (1938). Science, technology and society in seventeenth century England. *Osiris*, *4*, 360–632.

(1973). *The sociology of science* (N. W. Storer, ed.). University of Chicago Press.

Michelson, A. A. (1879). Experimental determination of the velocity of light. *Nature*, *21*, 94–96, 120–122, 226.

Mill, J. S. ([1859]1963). On liberty. In J. M. Robson (ed.), *Collected works: Essays on politics and society* (Vol. 18, pp. 213–310). University of Toronto Press.

Morgan, M., & Morrison, M. (1999). *Models as mediators: Perspectives on natural and social science*. Cambridge University Press.

Morrison, D. E., & Henkel, R. E. (eds.). (1970). *The significance test controversy: A reader*. Chicago, IL: Aldine Pub. Co.

Morrison, M. (2000). *Unifying scientific theories: Physical concepts and mathematical structures*. New York: Cambridge University Press.

Musgrave, A. (1974). Logical versus historical theories of confirmation. *The British Journal for the Philosophy of Science*, *25*(1), 1–23.

(1982). Constructive empiricism versus scientific realism. *The Philosophical Quarterly*, *128*, 262–271.

(2010). Critical rationalism, explanation, and severe tests. In D. G. Mayo & A. Spanos (eds.), *Error and inference: Recent exchanges on experimental reasoning, reliability, objectivity, and rationality* (pp. 88–112). New York: Cambridge University Press.

National Science Board. (2012). *Science and engineering indicators.* Retrieved April 16, 2013, from www.nsf.gov/statistics/seind12/start.htm

Neurath, O. (1938/1955). Unified science as encyclopedic integration. In O. Neurath, R. Carnap, & C. W. Morris (eds.), *International encyclopedia of unified science* (Vol. 1, pp. 1–27). University of Chicago Press.

Newton, I. (1978). *Isaac Newton's papers and letters on natural philosophy* (Second edn.; I. B. Cohen, ed.). Cambridge, MA: Harvard University Press.

([1730]1979). *Opticks* (Fourth ed.). New York: Dover Publications.

(1984). *The optical papers of Isaac Newton* (Vol. 1; A. E. Shapiro, ed.). Cambridge University Press.

([1687]1999). *The principia: Mathematical principles of natural philosophy* (I. B. Cohen & A. Whitman, eds. & trans.). Berkeley, CA: University of California Press.

Neyman, J. (1950). *First course in probability and statistics*. New York: Henry Holt and Company.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika*, *20A*(1/2), 175–240.

(1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *231*, 289–337.

Norton, J. D. (1993). General covariance and the foundations of general relativity: Eight decades of dispute. *Reports of Progress in Physics*, *56*, 791–861.

(2010). Cosmic confusions: Not supporting versus supporting not. *Philosophy of Science*, *77*, 501–523.

Okruhlik, K. (1994). Gender and the biological sciences. *Canadian Journal of Philosophy*, *20 (supplementary)*, 21–42.

Parker, W. (2009). Does matter really matter? Computer simulations, experiments, and materiality. *Synthese*, *169*, 483–496.

Pashby, T. (2012). Dirac's prediction of the positron: A case study for the current realism debate. *Perspectives on Science*, *20*(4), 440–475.

Peirce, C. S. (1883). A theory of probable inference. In C. S. Peirce (ed.), *Studies in logic: by members of the Johns Hopkins University* (pp. 126–181). Boston, MA: Little, Brown, and Company.

Perrin, J. (1916). *Atoms* (D. L. Hammick, trans.). New York: D. Van Nostrand.

Pinnick, C. L. (1994). Feminist epistemology: Implications for philosophy of science. *Philosophy of Science*, *61*(4), 646–657.

Pinnick, C. L., Koertge, N., & Almeder, R. (eds.). (2003). *Scrutinizing feminist epistemology: An examination of gender in science*. New Brunswick, NJ: Rutgers University Press.

Poincaré, H. ([1905]1952). *Science and hypothesis* (F. Maitland, trans.). Mineola, NY: Dover.

Popper, K. (1965). *Conjectures and refutations*. New York: Harper.

([1959]1992). *The logic of scientific discovery*. New York: Routledge.

Porter, T. (1988). *The rise of statistical thinking: 1820–1900*. Princeton University Press.

Pratt, J. W. (1977). 'Decisions' as statistical evidence and Birnbaum's 'confidence concept'. *Synthese*, *36*(1), 59–69.

Preston, T. (1890). *The theory of light*. New York: Macmillan and Co.

Psillos, S. (1999). *Scientific realism: How science tracks truth*. New York: Routledge.

Putnam, H. (1978). *Meaning and the moral sciences*. London: Routledge.

(1979). What is mathematical truth? In *Mathematics, matter, and method: Philosophical papers* (Vol. 1, pp. 60–78). New York: Cambridge University Press.

Quine, W. v. O. (1951). Two dogmas of empiricism. *The Philosophical Review*, *60*, 20–43.

(1980). *From a logical point of view* (Second edn.). Cambridge, MA: Harvard University Press.

(1991). Two dogmas in retrospect. *Canadian Journal of Philosophy*, *21*, 265–274.

Ramsey, F. P. (1950). Truth and probability. In R. B. Braithwaite (ed.), *The foundations of mathematics and other logical essays* (pp. 156–198). London: Routledge and Kegan Paul.

Reichenbach, H. (1938). *Experience and prediction*. University of Chicago Press.

Reisch, G. A. (1991). Did Kuhn kill logical empiricism? *Philosophy of Science*, *58*(2), 264–277.

Resnik, D. B. (1994). Hacking's experimental realism. *Canadian Journal of Philosophy*, *24*, 395–411.

Richards, R. (2002). Kuhnian values and cladistic parsimony. *Perspectives on Science*, *10*, 1–27.

Richardson, A., & Uebel, T. (eds.). (2007). *The Cambridge companion to logical empiricism*. New York: Cambridge University Press.

Roberts, J. T. (2008). *The law-governed universe*. New York: Oxford University Press.

Roqué, X. (1997). The manufacture of the positron. *Studies in History and Philosophy of Modern Physics*, *28*, 73–129.

Rudge, D. W. (1998). A Bayesian analysis of strategies in evolutionary biology. *Perspectives on Science*, *6*, 341–360.

(2001). Kettlewell from an error statistician's point of view. *Perspectives on Science*, *9*, 59–77.

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of Science*, *20*(1), 1–6.

Rudwick, M. (1985). *The meaning of fossils: Episodes in the history of paleontology* (Second edn.). University of Chicago Press.

Salmon, W. C. (1981). Rational prediction. *The British Journal for the Philosophy of Science*, *32*, 115–125.

(1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.

(1990). Rationality and objectivity in science or Tom Kuhn meets Tom Bayes. In C. W. Savage (ed.), *Scientific theories* (pp. 175–204). Minneapolis, MN: University of Minnesota Press.

(1994). Causality without counterfactuals. *Philosophy of Science*, *61*(2), 297–312.

Savage, L. (ed.). (1962). *The foundations of statistical inference: A discussion*. London: Methuen.

(1972). *Foundations of statistics* (Second edn.). New York: Dover.

Schweber, S. (1994). *QED and the men who made it*. Princeton University Press.

Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science*, *130*(3374), 477–482.

(1962). Explanations, predictions, and laws. In H. Feigl & G. Maxwell (eds.), *Minnesota studies in the philosophy of science* (Vol. 3: Scientific Explanation, Space, and Time, pp. 170–230). Minneapolis, MN: University of Minnesota Press.

Searle, J. (1969). *Speech acts: An esssay in the philosophy of knowledge*. Cambridge University Press.

Searle, J., & Vanderken, D. (1985). *Foundations of illocutionary logic*. Cambridge University Press.

Senn, S. (2001). Two cheers for p-values? *Journal of Epidemiology and Biostatistics*, *6*, 193–204.

(2011). You may believe you are a Bayesian but you are probably wrong. *Rationality, Markets, and Morals*, *2*, 48–66.

Shapere, D. (1964). The structure of scientific revolutions. *The Philosophical Review*, *73*(3), 383–394.

(1966). Meaning and scientific change. In R. G. Colodny (ed.), *Mind and cosmos: Essays in contemporary science and philosophy* (pp. 41–85). University of Pittsburgh Press.

(1993). Astronomy and anti-realism. *Philosophy of Science*, *60*, 134–150.

Shapiro, A. (2002). Newton's optics and atomism. In *The Cambridge companion to Newton.* New York: Cambridge University Press.

Sheehan, W., Kollerstrom, N., & Waff, C. B. (2004). The case of the pilfered planet: Did the British steal Neptune? *Scientific American*, *291*, 92–99.

Silver, N. (2012, July). *Measuring the effects of voter identification laws.* Blog post. Retrieved August 1, 2012, from http://fivethirtyeight.blogs.nytimes.com/2012/07/15/measuring-the-effects-of-voter-identification-laws/

Smith, L. A. (2007). *Chaos: A very short introduction*. New York: Oxford University Press.

Sneed, J. (1971). *The logical structure of mathematical physics*. Dordrecht: Reidel.

Snyder, L. (2006). *Reforming philosophy: A Victorian debate on science and society*. University of Chicago Press.

(2011). *The philosophical breakfast club: Four remarkable friends who transformed science and changed the world*. New York: Broadway Books.

Spanos, A. (1999). *Probability theory and statistical inference*. Cambridge University Press.

(2010). Is frequentist testing vulnerable to the base-rate fallacy? *Philosophy of Science*, *77*(4), 565–583.

Stadler, F. (ed.). (2003). *The Vienna Circle and logical empiricism*. Dordrecht: Kluwer.

Staley, K. W. (1999a). Golden events and statistics: What's wrong with Galison's image/logic distinction? *Perspectives on Science*, *7*(2), 196–230.

(1999b). Logic, liberty, and anarchy: Mill and Feyerabend on scientific method. *The Social Science Journal*, *36*, 603–614.

(2008). Error-statistical elimination of alternative hypotheses. *Synthese*, *163*, 397–408.

(2012). Strategies for securing evidence through model criticism. *European Journal for Philosophy of Science*, *2*, 21–43.

(2013). Experimental knowledge in the face of theoretical error. In M. Boumans, G. Hon, & A. Petersen (eds.), *Error and uncertainty in scientific practice.* London: Pickering and Chatto.

Staley, K. W., & Cobb, A. (2011). Internalist and externalist aspects of justification in scientific inquiry. *Synthese*, *182*, 475–492.

Standage, T. (2000). *The Neptune file: A story of mathematical cunning, astronomical rivalry, and the discovery of new worlds*. New York: Walker and Co.

Stanford, K. (2006). *Exceeding our grasp: Science, history, and the problem of unconceived alternatives*. New York: Oxford University Press.

(2010). Getting real: The hypothesis of organic fossil origins. *The Modern Schoolman*, *87*, 219–241.

Steel, D. (2010). Epistemic values and the argument from inductive risk. *Philosophy of Science*, *77*(1), 14–34.

Stegmüller, W. (1976). *The structure and dynamics of theories*. New York: Springer-Verlag.

Stein, H. (1989). Yes, but . . . : Some skeptical remarks on realism and anti-realism. *Dialectica*, *43*, 47–65.

Steno, N. ([1667]1958). *The earliest geological treatise* (A. Garboe, trans.). London: Macmillan and Co.

Stigler, S. (1990). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Belknap Press of Harvard University Press.

Strevens, M. (2007). Review of Woodward, *Making Things Happen*. *Philosophy and Phenomenological Research*, *74*(1), 233–249.

(2008). *Depth: An account of scientific explanation*. Cambridge, MA: Harvard University Press.

Suppe, F. (1977). The search for philosophic understanding of scientific theories. In F. Suppe (ed.), *The structure of scientific theories* (Second edn., pp. 3–241). Urbana, IL: University of Illinois Press.

Suppes, P. (1962). Models of data. In E. Nagel, P. Suppes, & A. Tarski (eds.), *Logic, methodology, and philosophy of science: Proceedings of the 1960 International Congress* (pp. 252–261). Stanford University Press.

(2002). *Representation and invariance of scientific structures*. Stanford, CA: CSLI Publications.

Teller, P. (1973). Conditionalisation and observation. *Synthese*, *26*, 218–258.

Thagard, P. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy*, *75*, 76–92.

Tobin, W. (2003). *The life and science of Léon Foucault: The man who proved the earth rotates*. Cambridge University Press.

Tukey, J. (1960). A survey of sampling from contaminated distributions. In I. Olkin (ed.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 448–485). Stanford University Press.

US Department of Health and Human Services Public Health Service. (2011). *Report on carcinogens* (Twelfth edn.; Tech. Rep.). US Department of Health and Human Services Public Health Service.

US Environmental Protection Agency. (2012). *Formaldehyde.* Retrieved September 5, 2012, from www.epa.gov/iris/subst/0419.htm

van Fraassen, B. (1972). A formal approach to the philosophy of science. In R. G. Colodny (ed.), *Paradigms and paradoxes.* University of Pittsburgh Press.

(1980). *The scientific image*. New York: Oxford University Press.

(2002). *The empirial stance*. New Haven: Yale University Press.

Weber, M. (2005). *Philosophy of experimental biology*. New York: Cambridge University Press.

Weinberg, S. (2012, July 13). Why the Higgs boson matters. *The New York Times*.

Whewell, W. (1847). *The philosophy of the inductive sciences: Founded upon their history*. London: John W. Parker.

Wilholt, T. (2009). Bias and values in scientific research. *Studies in History and Philosophy of Science*, *40*, 92–101.

Will, C. M. (1993a). *Theory and experiment in gravitational physics* (Second edn.). New York: Cambridge University Press.

(1993b). *Was Einstein right? putting general relativity to the test* (Second edn.). New York: Basic Books.

Williamson, J. (2010). *In defence of objective Bayesianism*. New York: Oxford University Press.

Winsberg, E. (2010). *Science in the age of computer simulation*. University of Chicago Press.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Worrall, J. (1985). Scientific discovery and theory-confirmation. In J. Pitt (ed.), *Change and progress in modern science* (pp. 301–332). Dordrecht: Reidel.

(1989a). Fresnel, Poisson and the white spot: The role of successful predictions in the acceptance of scientific theories. In D. Gooding, T. Pinch, & S. Schaffer (eds.), *The uses of experiment: Studies in the natural sciences* (pp. 137–157). New York: Cambridge University Press.

(1989b). Structural realism: The best of both worlds? *Dialectica*, *43*, 99–124.

(2008). Evidence and ethics in medicine. *Perspectives in Biology and Medicine*, *51*, 418–431.

Wray, K. B. (1999). A defense of Longino's social epistemology. *Philosophy of Science*, *66*, S538–S552.

(2011). *Kuhn's evolutionary social epistemology*. New York: Cambridge University Press.

Wylie, A., & Hankinson Nelson, L. (2007). Coming to terms with the values of science: Insights from feminist science studies scholarship. In H. Kincaid, J. Dupré, & A. Wylie (eds.), *Value-free science? Ideals and illusions* (pp. 58–86). New York: Oxford University Press.

Wylie, A., Potter, E., & Bauchspies, W. K. (2012). Feminist perspectives on science. In E. N. Zalta (ed.), *The Stanford encyclopedia of philosophy* (Fall 2012 ed.). http://plato.stanford.edu/archives/fall2012/entries/feminist-science/.

Zahar, E. (1973). Why did Einstein's programme supersede Lorentz's? (i). *The British Journal for the Philosophy of Science*, *24*(2), 95–123.

Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. Ann Arbor, MI: University of Michigan Press.

Index

abduction, *see* inference to the best explanation
acceptance of hypotheses, 24, 179, 236–238
Achinstein, Peter, 100, 189–192, 211–216, 228
action potential, 202
  Hodgkin-Huxley model of, 203–204, 208–210, 228
activities, 227
Adams, John Couch, 75
aether, 185, 195
aims of science, 128, 179, 180, 242
  institutional, 244
alternative hypothesis, 148
analytic statements, 34
Anderson, Carl, 169, 192
androcentrism, 251, 252
anomaly, 57, 75, 77
anti-realism, 48, 67
"anything goes", 87
Arago, Françoise, 28, 75, 195
argument from inductive risk, 233–240, 242
Aristotelianism, 94, 96, 121
Aristotle, 93, 217
  on explanation, 199
Avogadro's number, 190

Barnes, Eric, 219
Batterman, Robert, 222n
Bayes's theorem, 106, 118–120, 124, 130
Bayes, Rev. Thomas, 106
Bayesianism, 111–134, 135, 164, 191n, 237
  personalist, 114, 132
behavioristic construal of statistical tests, 151–152, 156, 160, 238–239, 240
Bernard, Claude, 210
betting rate, 113
  fair, 113, 117
binomial distribution, 138, 147
Birnbaum, Allan, 152
Blackett, P.M.S., 170, 192
Bohr, Niels, 58
Brown, Ernest William, 80
Brownian motion, 190

calibration, 37
Carnap, Rudolf, 43, 44, 132, 229n
Cartwright, Nancy, 188, 256n
catch-all hypothesis, 123, 126–128, 133
chance-based odds, 117
chaos, 219n
Charles' Law, 46
Chevalier's problem, 107, 134
Churchman, C. West, 238, 240
circularity
  of appeal to paradigms, 60, 69
  of Woodward's manipulationist account, 226
  vicious- vs. premise-, 173
cloud chamber, 169n, 169
coherence, 114, 120
communism, 245
conditionals, 16, 43
  counterfactual, 224
consilience of inductions, 80n
content noun, 213
conventions, 22–23

Copernicus, Nicolaus, 93
correspondence rules, 42, 45, 50
corroboration, 19–20
counterinduction, 89–97
Cox, David, 132
Craver, Carl, 204, 209–210, 227
crisis (Kuhnian), 58
curare poisoning, 210–211, 225, 226

D-N model, 199
Darden, Lindley, 227
Darwin, Charles, 90
  evolutionary theory of, 67, 68, 217
demarcation of science from non-science, 20, 23
determinism, 219n, 219
devil's advocate, 85, 91, 99
Dirac, Paul, 168–169, 196
disinterestedness, 245
domain of discourse, 51
Douglas, Heather, 240–243
Dowe, Phil, 221n
Duhem's problem, *see* underdetermination, problem of
Duhem, Pierre, 27–28, 30–33, 39, 190, 194, 199n, 237
Dutch book, 114n, 114

Einstein, Albert, 20, 41, 59
  dynamics of, 62
electromagnetism
  Maxwell's theory of, 195
Elliott, Kevin, 243
empirical adequacy, 179
empirical basis, 24
empiricism, 168, 178, 191
  constructive, 179–183
  logical, 33, 39–44, 48, 55, 56, 62, 69, 71, 96, 105, 132, 178, 216, 229n, 229
epidemiological studies, 157, 163
epistemological anarchism, 84, 86–102
error
  ability of test to reveal, 153, 156, 164
  arguments from, 156
  ruling out, 11–14, 37
  type I vs. type II, 149, 235–236
error-statistical philosophy, 135, 153–166, 255, 256
evidence, 161, 191, 209
  contextual account of, 246
  Mayo's minimal principle for, 153
  securing of, 256
  statistical, 152
evolutionary biology, 57, 68
expectation value, 138, 142
experimentation, 39, 187
  coherence of, 188
  epistemology of, 37, 38
explanandum, 184, 200
explanans, 184, 200
explanation, 199–228
  Achinstein's pragmatic theory, 211–216
  asymptotic, 222n
  causal, 209, 220–228
  causal mechanical account, 221–222
  causal, manipulationist account, 223–226
  covering law model, 199–210, 215, 228
  mechanistic, 226–228
  ordered pair account, 213–214, 216
  unificationist theories, 216–220
explanation sketch, 201, 205
explanatory models, 247
explanatory store, 218
extension (of predicate), 52

fallibilism, 89
falsifiability, 20
falsificationism, 15–25, 26, 73–74, 77, 83, 89
  dogmatic, 73
  methodological, 73
  naïve vs. sophisticated methodological, 74, 76, 89n
Faraday, Michael, 69
feminist critiques, 251–254
Feyerabend, Paul, 71, 84–102, 252
filling instructions, 218
Fine, Arthur, 198
Fisher, Ronald, 142, 144, 162
Fizeau, Hippolyte, 28
Fizeau-Foucault experiment, 28–30, 36

formaldehyde as carcinogen, 157–163, 234–236
fossils, 121–127
Foucault, Léon, 28, 36
Franklin, Allan, 37
fraud, 245, 245n, 245
free will, xiii
frequentism, *see* probability, as relative frequency
Fresnel, Augustin Jean, 194
Freudian psychoanalytic theory, 21
Friedman, Michael, 216n
function, 52n
funding of science, 231–232
  and bias, 240

Galilei, Galileo, 92–97
Glennan, Stuart, 227
"good sense", 32, 38, 64
Grene, Marjorie, 249
gun violence research, 232

Hacking, Ian, 187–189, 192
Hall, Asaph, 79
Hanson, Norwood Russell, 61n, 77
Harding, Sandra, 253
Harper, William, 79
Hartsock, Nancy, 253
Heisenberg, Werner, 59
Hempel, Carl, 199–202, 204, 216
Hesse, Mary, 69
heterosexism, 251
Higgs boson, discovery of, 141, 143
hole theory (Dirac), 168–169, 177, 186, 196
Hooke, Robert, 13
Hume, David, 221n
Huygens, Christian, 13

IID (independent and identically distributed), 138, 147
illocutionary
  acts, 211
  force, problem of, 212
incommensurability, 61–63, 67, 68
  methodological, 61
  semantic, 62
induction, 90
  problem of, 5–14, 15, 26
inductive-statistical model, 201
inductivism, 15, 26, 83
inertia, principle of, 95
inference to the best explanation, 172–174, 183, 189, 193
instrumentalism, xxi, 171
Intemann, Kristen, 254
intervention, 224–225

Jeffrey, Richard, 237

Kitcher, Philip, 216–219
knowledge, 249n
Kuhn, Thomas, 55–70, 71, 73, 93n, 125, 195

Lakatos, Imre, 57, 71–84, 184
Large Hadron Collider, 141
Laudan, Larry, 35, 175–178
laws, 200, 204–206, 206n
  of equivalence, 207
Le Verrier, Urbain, 75–78
Leplin, Jarrett, 184
Levi, Isaac, 238, 240
Lewis, Peter, 178
linear hormonal model, 247–248
Lister, Martin, 125
Lloyd, Elisabeth, 99
Longino, Helen, 246–251, 252–254

Mach, Ernst, 190
Machamer, Peter, 227
MacIntyre, Alasdair, 249
man-the-hunter approach, 252
manipulation, 223, 224
Marxist theory of history, 21
Massimi, Michela, 189, 192
Maxwell, Grover, 180
Mayo, Deborah, 132, 151n, 153–166, 239, 255
McMullin, Ernan, 198
mechanism, *see* explanation, mechanistic
Mendel, Gregor, 68
Mercury, anomalous orbit of, 77
Merton, Robert K., 243n, 244–246

metaphysics, 41, 43, 197, 199n
Mill's methods, *see* Mill, John Stuart, *A System of Logic*
Mill, John Stuart, 87, 101
  A System of Logic, 101
  On Liberty, 97–99
Millikan, Robert A., 169n
models
  adequacy of, 147
  idealization in, 46–48, 137
  in formal semantics, 54
  in semantic view, 45–48
  probability, 108
modus tollens, 18, 23, 26
Musgrave, Alan, 81

natural interpretations, 95
natural ontological attitude (NOA), 198
natural philosophy, xiii
Neoplatonism, 121
Neurath, Otto, 55
New Experimentalism, 186
Newcomb, Simon, 77, 79
Newton, Isaac
  celestial mechanics of, 75, 76, 220
  dynamics of, 62, 195
  emission theory of light, 28, 30
  laws of motion, 47
  optical experiments, 8–11, 156
  rules for study of natural philosophy, 90
  theory of gravity, 69, 79, 217
Neyman, Jerzy, 147, 151n, 162
Neyman-Pearson tests, 146–151, 154n
  power of, 149–150
  significance level of, 149–150
no miracles argument, 171, 175–177, 194
  circularity of, 173–174
nomic expectability, 201, 206–207
normal science, 56–57, 70
norms of science (Merton), 243–246
novel prediction, 79–82, 184
  heuristic, 82, 184
  temporal, 80
null hypothesis, 142, 144, 146, 234
  null distribution as model of, 142, 143, 147

objectivity, 132, 243, 246, 248–249, 254
  of scientific community, 250–251
  strong, 253
observability, 180–182, 191
observation
  theory-ladenness of, 62
Occhialini, G.P.S., 170, 192
Oppenheim, Paul, 201
organized skepticism, 246
Ostwald, Wilhelm, 190

p-value, 143n, 143–146, 159
paradigm, 56–63, 64
  as disciplinary matrix, 57, 63
  as exemplar, 57, 66
Parametrized Post-Newtonian (PPN) framework, 165n
paresis, 206
Pascal, Blaise, 107
Pashby, Thomas, 186, 196
Pauli, Wolfgang, 59
Pearson, Egon, 147, 151n
Peirce, Charles S., 154n, 172n
Perrin, Jean, 190–192
pessimistic meta-induction, 177–178
phylogenetics, 64–67, 68
  molecular, 57, 66–67
Pinnick, Cassandra, 253
pluralism, 251–254, *see* proliferation
Poincaré, Henri, 190, 194, 195
Poisson, Siméon Denis, 194
Popper, Karl, 15–25, 26, 36, 71, 72
positivism, *see* empiricism, logical
positron, 183
  as experimental tool, 187, 189, 193n, 193
  discovery of, 169–171, 192–193
  prediction of, 168
pragmatism, 154n
predicate, 50
prediction, 201, 204, 209
primates as model organisms, 232–233
probability
  as degree of belief, 111–134
  as relative frequency, 111, 135–166
  axioms, 109
  conditional, 119

function, 109, 110, 119
posterior, 119, 125, 126, 132
prior, 119, 123, 126, 127
theory, 105
progress of science, 67
proliferation, 90, 100
Psillos, Stathis, 173, 184, 185
Ptolemy, Claudius, 93
Putnam, Hilary, 171, 177

quantifier, 51
quantum electrodynamics, 177, 196
quantum mechanics, 59
"Old Quantum Theory", 58–59
Quine, W.V.O., 33

Ramsey, Frank P., 116
random variable, 107
rational reconstruction, 83
rationality, 83, 114, 128, 130
realism
anti-, 167–168, 171, 191, 194
divide and conquer strategy, 185–186, 195
entity, 187–189,
epistemic structural, 194–196
experimental argument for, 189–192
ontic structural, 194, 197
scientific, 48, 167, 171–174, 184–186
reduction sentences, 42, 43
reference priors, 131
Reichenbach, Hans, 136, 140
relativity
general theory of, 20, 41, 78, 165
of motion, Galileo's principle of, 96
special theory of, 168, 195
research programs
hard core of, 74, 76, 79
methodology of scientific, 74–84
negative heuristic, 75, 76
positive heuristic, 75, 77, 79
progressive vs. degenerating, 78
protective belt of, 74, 76
Resnik, David, 188
Retraction Watch, 245n
revolutionary science, *see* scientific revolutions
Rudner, Richard, 236–238

σ, *see* standard deviation
Salmon, Wesley, 25, 190, 221
satisfaction, 51
scientific progress, 82
scientific revolutions, 57, 59–61, 63, 69, 194
as challenge to scientific realism, 174–178
Scriven, Michael, 205
security, *see* evidence, securing of
selectionist model, 247–248
semantics, 49
sentence schema, 50
severity, 154–157, 193
analysis, 160–163, 165
principle, 154, 160, 164
sexism, 251
size of statistical test, *see* Neyman-Pearson tests, significance level of
Snyder, Laura, 99
sociology of science, 243n
soundness, 4
stake, 113
standard deviation, 138, 141, 142
standpoint epistemology
feminist, 253–254
Marxist, 253
Stanford, Kyle, 198
statistical significance, 140–146, 159
observed, 143
statistics, 37, 69
Bayesian, 131–132
Steel, Daniel, 241
Stein, Howard, 198
Steno (Niels Stensen), 122–123
structural unemployment, 222
structure, 51
realism about, *see* realism, epistemic structural and ontic structural
subjective Bayesianism, *see* Bayesianism, personalist
success, 184
sure-loss contract, 114–115, 120

synthetic statements, 34, 44
systematics, 64
systematization, 218

test statistic, 142
theory structure
  semantic view of, 45–48, 179
  syntactic view of, 44, 48
transformative criticism, 250

underdetermination, 26–38, 73, 182–183, 189, 209, 248, 254
  Kuhnian, 63
  problem of, 32–33, 127n
  Quine's thesis of, 34
  thesis, 26
understanding, 199, 212, 213
unified science, 55
uniformity of nature, principle of, 7
universalism, 244
updating by conditionalization, 120, 124, 130
Uranus, anomalous orbit of, 75
use novelty, *see* novel prediction, heuristic

vagueness, 181
validity, deductive, 4
value-free ideal, 229, 246, 252, 254
  as autonomy, 230, 233
  as impartiality, 230, 255
  as neutrality, 230, 255
values, 63, 229–257
  cognitive, 230n
  constitutive, 246
  contextual, 246
  direct vs. indirect roles, 240–243
  epistemic, 230, 241
  intellectual, 73, 82
van Fraassen, Bas, 179–183
verification principle, 41–44
Vienna circle, 40–41, 49, 50, 55
Vulcan, 78

wagers, 113
wave theory of light, 13, 185, 194
Weber, Marcel, 209n
Whewell, William, 80n
Williamson, Jon, 132
Wilson, C.T.R., 169n
woman-the-gatherer approach, 252
Woodward, James, 220, 223–226
Worrall, John, 184, 194
wouralia, *see* curare poisoning
Wray, K. Brad, 68